The Battle of the Neighborhoods

Introduction

Business Problem

In this project we will try to find an optimal location for a **yoga studio**. Specifically, this report will be targeted to stakeholders interested in opening a yoga studio in London.

Since there are lots of yoga studios in London we will try to detect locations that are not already crowded with yoga studios and gyms (since many gyms provide group classes on yoga). We are also particularly interested in areas with no yoga studio in vicinity. We would also prefer locations as close to city center 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.

Stakeholders interested in this project

Business man interested in opening a yoga studio in London, not far from city center and probably close to main business district – The City of London, if possible.

Data

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

- number of existing yoga studios and gyms in the neighborhood
- number of and distance to yoga studios in the neighborhood, if any
- distance of neighborhood from Trafalgar Square

We decided to use regularly spaced grid of locations, centered around city center, 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 yoga studios and their type and location in every neighborhood will be obtained using **Foursquare API**
- coordinate of The City of London center will be obtained using **Google Maps API geocoding** of well-known location (Trafalgar Square)

Methodology

In this project we directed our efforts on detecting areas of London that have low gym density, particularly those with low number of yoga studios. We limited our analysis to area ~6km around city center.

In first step we have collected the required data: location and type (category) of every gym within 6km from London center (Trafalgar Square). We have also identified yoga studios (according to Foursquare categorization).

Second step in our analysis was calculation and exploration of 'gym density' across different areas of London - we used heatmaps to identify a few promising areas close to center with low number of gyms in general (and no yoga studio in vicinity) and focus our attention on those areas.

In third and final step we focused on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we took into consideration locations with no more than two gyms in radius of 250 meters, and we wanted locations without yoga studios in radius of 400 meters. We presented map of all such locations but also created 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.

Results

Our analysis showed that although there is a great number of gyms in London (~2000 in our initial area of interest which was 12x12km around Trafalgar Square), there are pockets of low gym density fairly close to city center. Highest concentration of gyms was detected north and west from Trafalgar Square, so we focused our attention to areas south-east and east, corresponding to borough - The City of London.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Trafalgar Square) we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two gyms in radius of 250m and those with an yoga studio 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.

Discussions

Result of all this is 15 zones containing largest number of potential new gym locations based on number of and distance to existing venues - both gyms in general and yoga studios particularly. This, of course, does not imply that those zones are actually optimal locations for a new studio! Purpose of this analysis was to only provide info on areas close to London center but not crowded with existing gyms (particularly yoga studios) - it is entirely possible that there is a very good reason for small number of gyms in any of those areas, reasons which would make them unsuitable for a new studio 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 center with low number of gyms (particularly yoga studios) in order to aid stakeholders in narrowing down the search for optimal location for a new yoga studio. By calculating gym density distribution from Foursquare data we have first identified general boroughs that justify further analysis (The City of London), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby gyms. 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 yoga studio 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, real estate availability, prices, social and economic dynamics of every neighborhood etc.