

Predicting the location for educational center

Arailym Akimova

Introduction: Business Problem

In this project I will try to find an optimal location for an educational center. It is a real task for, because We are with my friends are going to open a school that will educate children on programming, robotics.

Since there are many educational centers, schools we will try to detect locations that are not crowded with educational centers but crowded with public schools that are not specialized in robotics.

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.

Data

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

- number of existing educational centers in the neighborhood
- number of existing public schools in the neighborhood
- distance of neighborhood from city center

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 centers and their type and location in every neighborhood will be obtained using Foursquare API
- coordinate of London center will be obtained using Google Maps API geocoding of well-known London location

Methodology

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

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

Second step in our analysis will be calculation and exploration of 'educational center density' across different areas of London - we will use heatmaps to identify a few promising areas close to center with low number of educational centers in general (and no educational centers 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 educational centers in radius of 250 meters, and we want locations without educational centers 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 / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Results

I

Our analysis shows that although there is a great number of educational centers in London (~2000 in our initial area of interest which was 12x12km around center), there are pockets of low educational center density fairly close to city center. Highest concentration of educational centers was detected north and west from Necessary district, so we focused our attention to areas south, south-east and east, corresponding to boroughs Kreuzberg, Friedrichshain and south-east corner of central Mitte borough. Another borough was identified as potentially interesting (Prenzlauer Berg, north-east from Necessary district), but our attention was focused on Kreuzberg and Friedrichshain which offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics *and* a number of pockets of low educational center density.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Necessary district) we first created a dense grid of location candidates (spaced 100m apart); those locations were then filtered so that those with more than two educational centers in radius of 250m and those with an Programming educational center 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 15 zones containing largest number of potential new educational center locations based on number of and distance to existing venues - both educational centers in general and Programming educational centers particularly. This, of course, does not imply that those zones are actually optimal locations for a new educational center! Purpose of this

analysis was to only provide info on areas close to London center but not crowded with existing educational centers (particularly Programming) - it is entirely possible that there is a very good reason for small number of educational centers in any of those areas, reasons which would make them unsuitable for a new educational center 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 educational centers (particularly programming educational centers) in order to aid stakeholders in narrowing down the search for optimal location for a new Programming educational center. By calculating educational center 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 educational centers. 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 educational center 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.