

Capstone Project - The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera

Table of contents

- [Introduction: Business Problem](#)
- [Data](#)
- [Methodology](#)
- [Analysis](#)
 - [Obtain Coordinates, Boundaries, Population Density of London Boroughs](#)
 - [Generate Neighborhoods](#)
 - [Venues](#)
 - [Neighborhoods Clustering](#)
 - ['Candidate' Neighborhoods](#)
- [Results and Discussion](#)
- [Conclusion](#)

Introduction: Business Problem

John with his family lives in London. He really likes his house and the neighborhood. However kids are growing and the current house seems too small for them so John is thinking of buying a bigger house.

Unfortunately there are not that many offers on the real estate market in his neighborhood. John has been checking adverts for several months but couldn't find anything suitable. So John decided to have a look at the houses available in the other areas of London.

This project will help John to narrow down his search and offer him the list of 'candidate' neighborhoods that would be very similar to the current 'reference' neighborhood in terms of the number and categories of venues around and the population density.

Methodology

For this project we'll generate a regularly spaced grid of locations covering the whole territory of London to define 'candidate' neighborhoods, then obtain information on the venues in each area using Foursquare API and use K-means clustering algorithm to find all neighborhoods similar to the 'reference' one.

Data

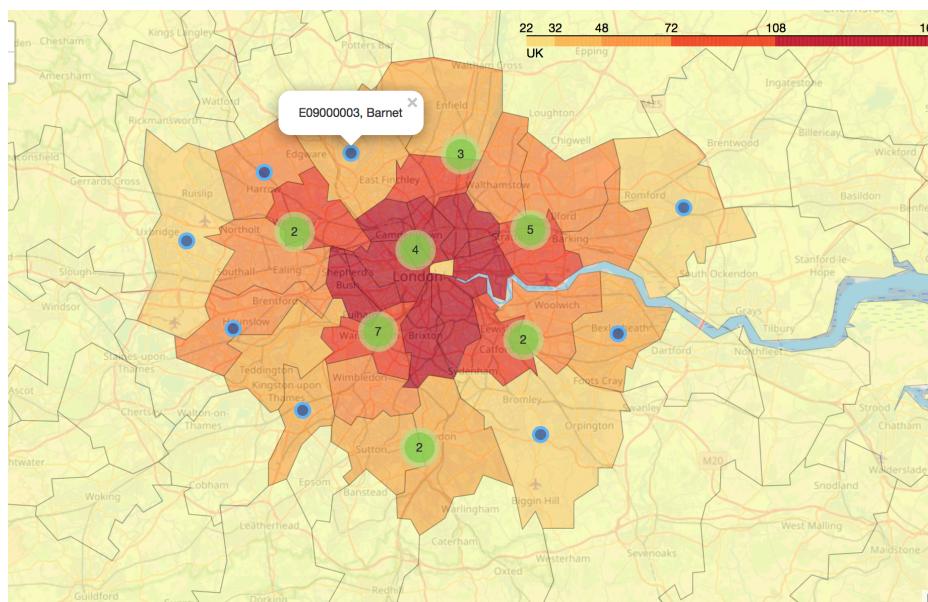
Following data sources will be needed to complete the project:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Nominatim library of the geopy package**
- number of venues and their type and location in every neighborhood will be obtained using **Foursquare API**
- coordinates of the centers and the boundaries of London boroughs will be downloaded from **the Open Geography Portal by the Office for National Statistics**, the UK's largest independent producer of official statistics and its recognised national statistical institute.

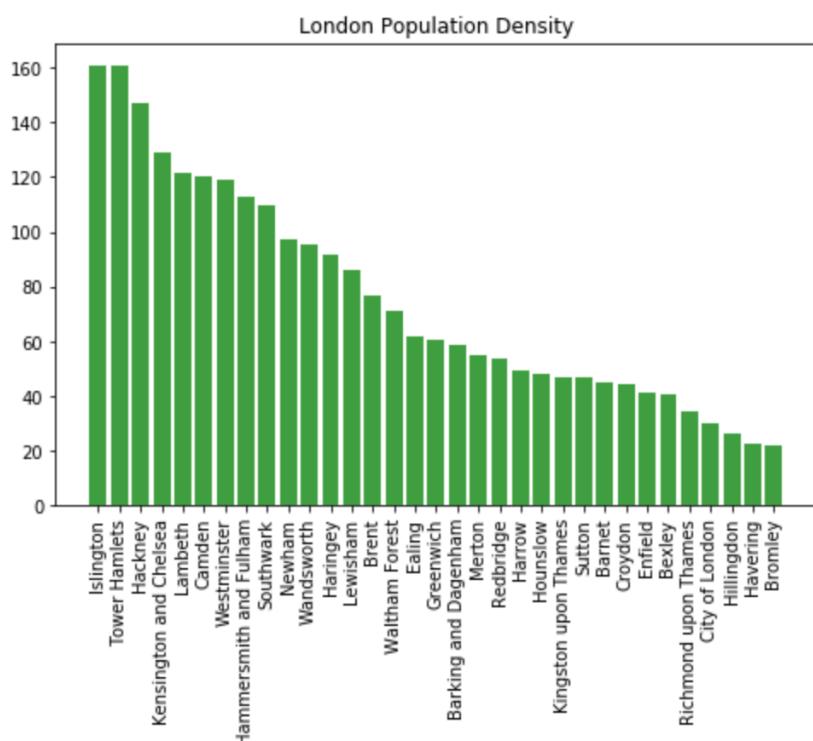
Analysis

Obtain Coordinates, Boundaries, Population Density of the London Boroughs

As the first step of the project we obtained information on the coordinates of the London boroughs from the Open Geography Portal by the Office for National Statistics. We downloaded the information on the squares of these areas from the same website. Then we got estimates of the population for the UK, England and Wales, Scotland and Northern Ireland from the Office for National Statistics website. Extracted information on the London boroughs from this data set and then calculated the population density for each London borough. The results were displayed on the choropleth map using the Folium package.



POPULATION DENSITY OF LONDON BOROUGHHS



POPULATION DENSITY OF LONDON BOROUGHHS

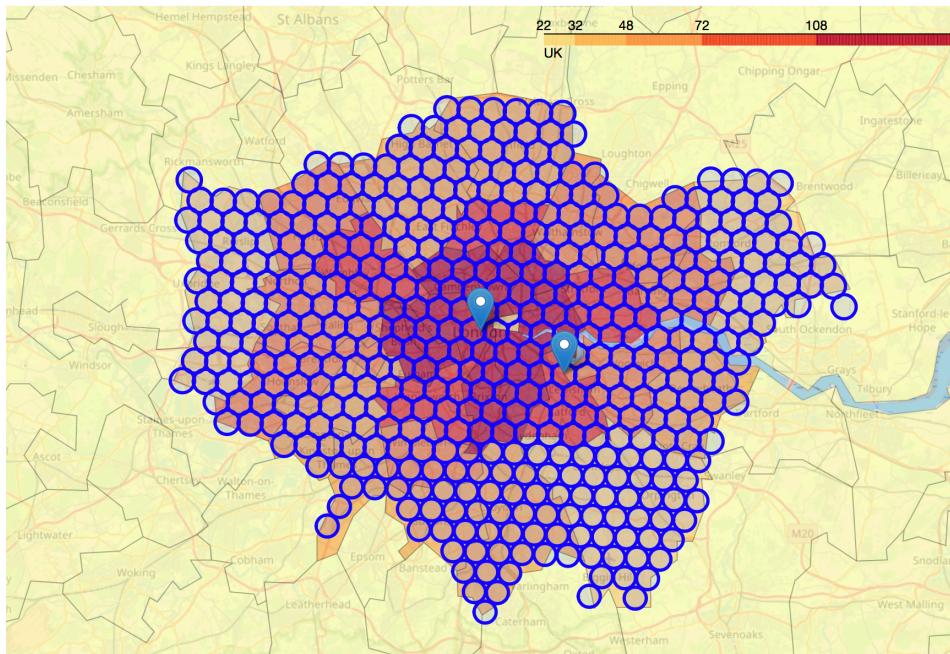
London has 33 boroughs. The most populated once are so called inner boroughs: Islington, Tower Hamlets, Hackney, etc. The least populated are so called outer boroughs: Hillingdon, Havering, Bromley.

Generate Neighbourhoods

On the next step of the project we generate the grid of the neighbourhoods that covered the whole territory of London.

Then we obtained addresses of the renters of all these locations suing Nominatim library of the geopy package.

464 ares were generated.



GRID OF NEIGHBOURHOODS

Venues

To obtain information on the number of venues and their categories in each neighbourhood we used Foursquare API.

13833 venues were found for all neighbourhoods.

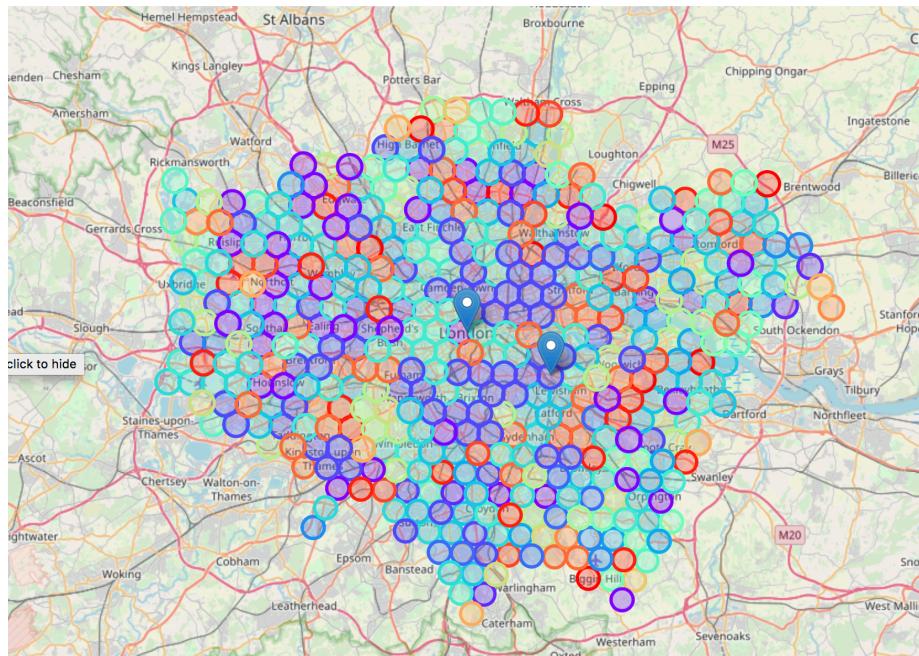
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Piles Wood, The Bridlepath, Chaldon, London Bo...	51.289188	-0.121330	Surrey National Golf course	51.286621	-0.109825	Golf Course
1	Piles Wood, The Bridlepath, Chaldon, London Bo...	51.289188	-0.121330	Horne Clay Shoot: Court Farm	51.285097	-0.123583	Athletics & Sports
2	Piles Wood, The Bridlepath, Chaldon, London Bo...	51.289188	-0.121330	Surrey Contracts Ltd	51.281850	-0.121909	Home Service
3	Farthing Downs Tree Belt, Downs Road, Purley, ...	51.305095	-0.134684	The Tudor Rose	51.306235	-0.123648	Restaurant
4	Farthing Downs Tree Belt, Downs Road, Purley, ...	51.305095	-0.134684	The Covered Way	51.303465	-0.143145	Moving Target

VENUES IN LONDON

Neighbourhoods Clustering

To cluster the neighbourhoods we used K-means algorithm. To obtain the optimum number of clusters we used the Elbow Method.

The map below shows the neighbourhoods split between 22 clusters.

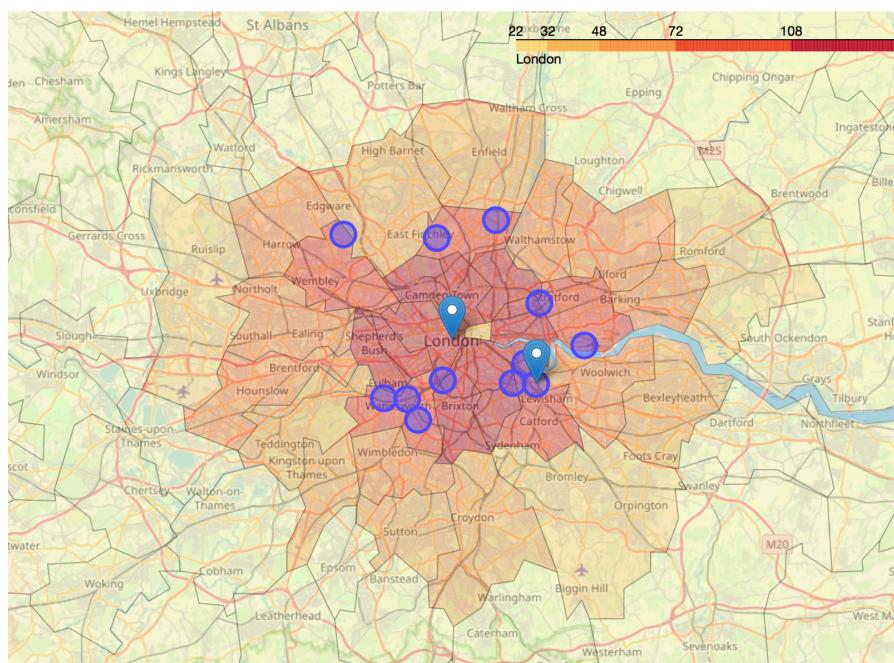


CLUSTERED NEIGHBOURHOODS

'Candidate' Neighbourhoods

The results were filtered to get all locations belong to the same cluster as the 'reference' one and then using one more condition - the population density range - the same as for the 'reference' neighborhood.

The algorithms produces 12 'candidate' neighbourhoods that could be further considered by John as the locations for his new home.



CANDIDATE NEIGHBOURHOODS

	Address	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Postcode	Borough
11	113, Pomeroy Street, Telegraph Hill, New Cross...	Store	Restaurant	Pub	Parks & Nature Object	Fast Food	SE14 5FY	Lewisham
12	Argosy House, Windlass Place, Hughesfield Esta...	Store	Restaurant	Parks & Nature Object	Pub	Transport	SE8 3QZ	Lewisham
22	The Crescent, 2, Brookmill Road, Hughesfield E...	Restaurant	Store	Pub	Bar	Sports Venue	SE8 4HJ	Lewisham
102	North Woolwich Road, Silvertown, London Boroug...	Hotel	Transport	Restaurant	Store	Fast Food	E16 2BB	Newham
107	Victoria Walk, Stratford Marsh, London Borough...	Entertainment	Store	Sports Venue	Café	Restaurant	E20 2ST	Newham
237	5, Oxford Road, Southfields, London Borough of...	Restaurant	Store	Pub	Café	Fast Food	SW15 2LG	Wandsworth
238	61,62,63, Nantes Close, Battersea, London Boro...	Store	Restaurant	Pub	Café	Fast Food	SW18 1JL	Wandsworth
240	Mills House, Thessaly Road, Patmore Estate, Ni...	Store	Restaurant	Pub	Café	Parks & Nature Object	SW8	Wandsworth
244	Trinity Road, Upper Tooting, London Borough of...	Pub	Store	Restaurant	Café	Sports Venue	SW18 3SH	Wandsworth
321	Hay Lane, Colindale, London Borough of Brent, ...	Store	Restaurant	Parks & Nature Object	Fast Food	Pub	NW9 0NH	Brent
453	Lordship Lane, Tottenham Hale, Tottenham, Lond...	Store	Restaurant	Pub	Transport	Bar	N17 8NS	Haringey
454	Woodland Rise, Shepherd's Hill, Muswell Hill, ...	Store	Restaurant	Café	Pub	Parks & Nature Object	N10 3UP	Haringey

CANDIDATE NEIGHBOURHOODS

Results and Discussion

To help John to find an ideal location for his new home we generated a list of the London areas which are very similar to the location he lives in right now and which he likes a lot. To achieve this:

- We obtained information on the coordinates of the London boroughs from the Open Geography Portal by the Office for National Statistics.
- We downloaded the information on the squares of these areas from the same website.
- Then we got estimates of the population for the UK, England and Wales, Scotland and Northern Ireland from the Office for National Statistics website. Extracted information on the London boroughs from this data set and then calculated the population density for each London borough.
- The results were displayed on the choropleth map using the Folium package.
- On the next step we generated the grid of areas with the 1000m radius covering the whole territory of London and obtained the information on the venues in each location using Foursquare API.
- We used K-means algorithm to cluster all the locations by the number and categories of venues they got.
- The results were filtered to get all areas in the same cluster as the 'reference' one and then using one more condition - the population density range - the same as for the 'ethalon' neighborhood.
- As the final step all 'candidate' neighborhoods were displayed on the map.

This project helped us to get the list of 12 'candidate' neighborhoods that John could consider as the new 'perfect' areas when looking for a new location of his new house.

Conclusion ¶

Purpose of this project was to identify London areas similar to the neighborhood where John lives now to help him to find a new location for his new house. To solve this problem we generate a grid of neighborhoods covering the whole territory of London.

Using Foursquare API we obtain information on the venues in each neighborhood - their number and types.

Then we clustered all areas with K-means algorithm to group similar areas into clusters.

We came up with the list of **75 'candidate' neighborhoods which John could consider to move in**.

There are lots of different official open data sources available that could be used to narrow down the search even further and help John to make a better and more informed decision. This could be **the house prices, school ratings, hospital ratings, unemployment rate, crime rate, etc.** The methodology used in this notebook could be refined further to incorporate the above listed data sources into the analysis and help to automate the decision making process.