Capstone Project - The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera

By Georgi Dezhov

Introduction: Business Problem

- In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an Indian restaurant in London, UK.

- Since there are lots of restaurants in London, we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Indian restaurants in vicinity. We would also prefer locations as close to city centre as possible, assuming that first two conditions are met.

- We will use our data science powers to generate a few most promising neighbourhoods 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 restaurants in the neighbourhood (any type of restaurant)
- number of and distance to Indian restaurants in the neighbourhood if any
- distance of neighbourhood from city centre

Data

We decided to use regularly spaced grid of locations, cantered 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 **Foursquare 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 **Foursquare API geocoding** of well-known London location (Trafalgar Square).

Neighbourhood Candidates

- Let us create latitude & longitude coordinates for centroids of our candidate neighbourhoods. We will create a grid of cells covering our area of interest which is approx. 12x12 kilometres cantered around London city centre.

- Let us first find the latitude and longitude of London city centre, using specific, well known address and Foursquare API geocoding.

The latitude and longitude of London city centre

- According to "http://shadyoldlady.com/location.php?loc=463" the exact centre of London is marked by a plaque in the Church of St. Martin's-in-the-Fields overlooking Trafalgar Square. So, we are going to look for the coordinates of Trafalgar Square.

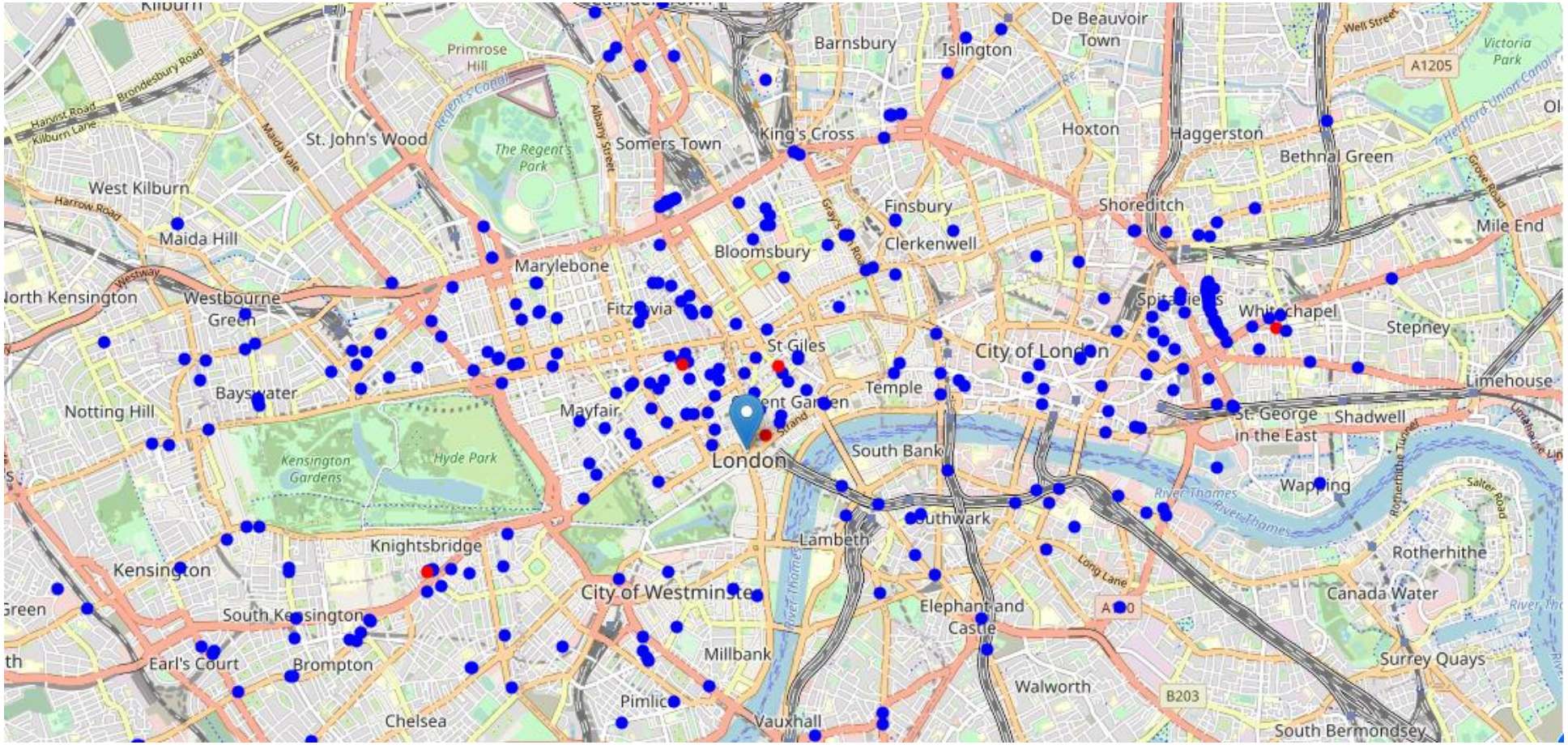
- The geographical coordinate of the London City center are [51.508037, -0.12804941070724718].

Foursquare Data

We are interested in venues in 'food' category, but only those that are proper restaurants - coffee shops, pizza places, bakeries etc. are not direct competitors so we do not care about those. So we will include in our list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific 'Indian restaurant' category, as we need info on Indian restaurants in the neighbourhood. We found:

- Total number of restaurants: 377
- Total number of Indian restaurants: 6
- Percentage of Indian restaurants: 1.59%
- Average number of restaurants in neighborhood: 0.9725274725274725

The collected restaurants in our area of interest and Indian restaurants in red color.



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 Indian 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 centre** (Trafalgar Square). We have also **identified Indian restaurants** (according to Foursquare categorization).

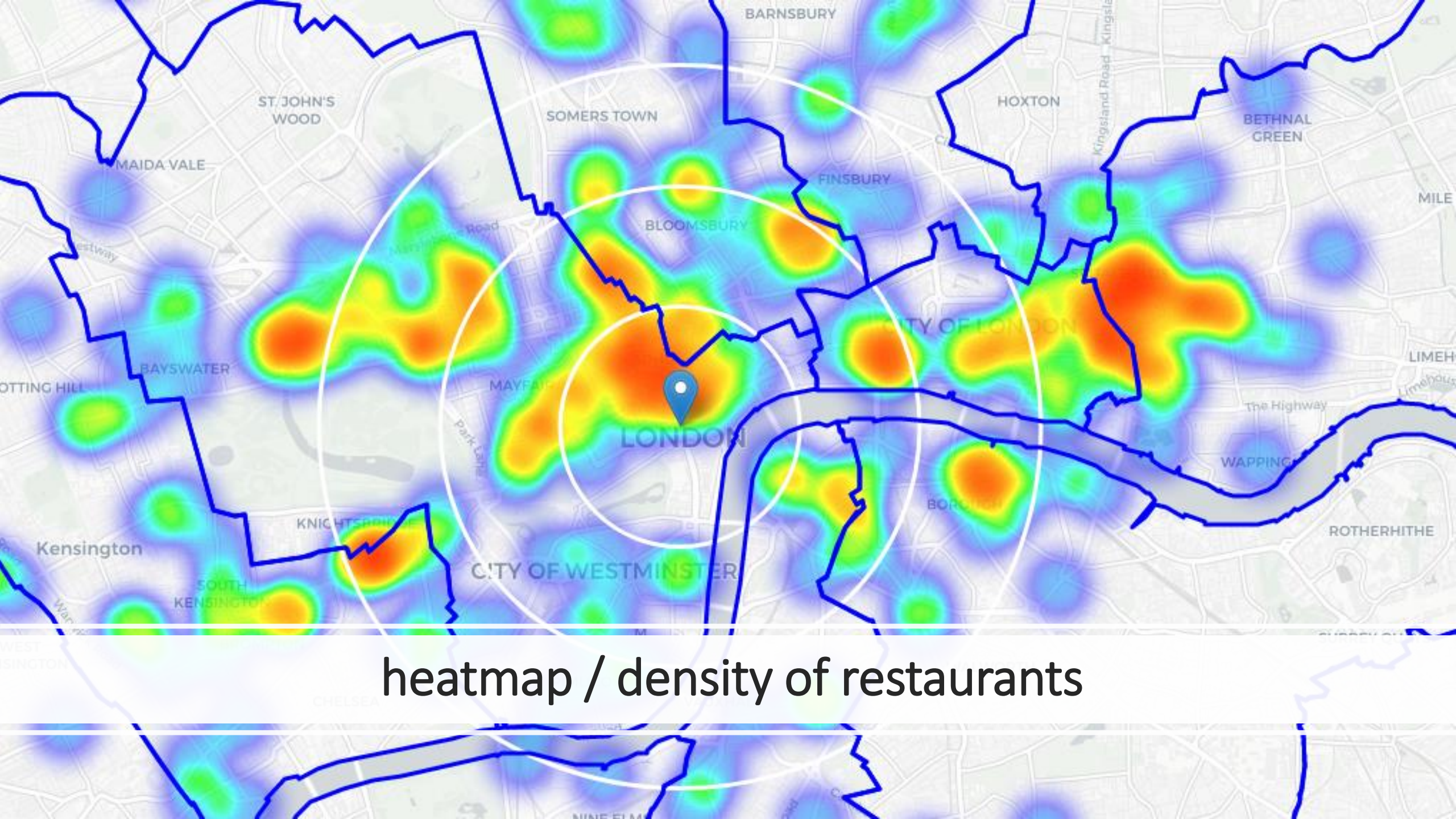
Methodology

- 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 Indian restaurants in vicinity) and focus our attention on those areas.

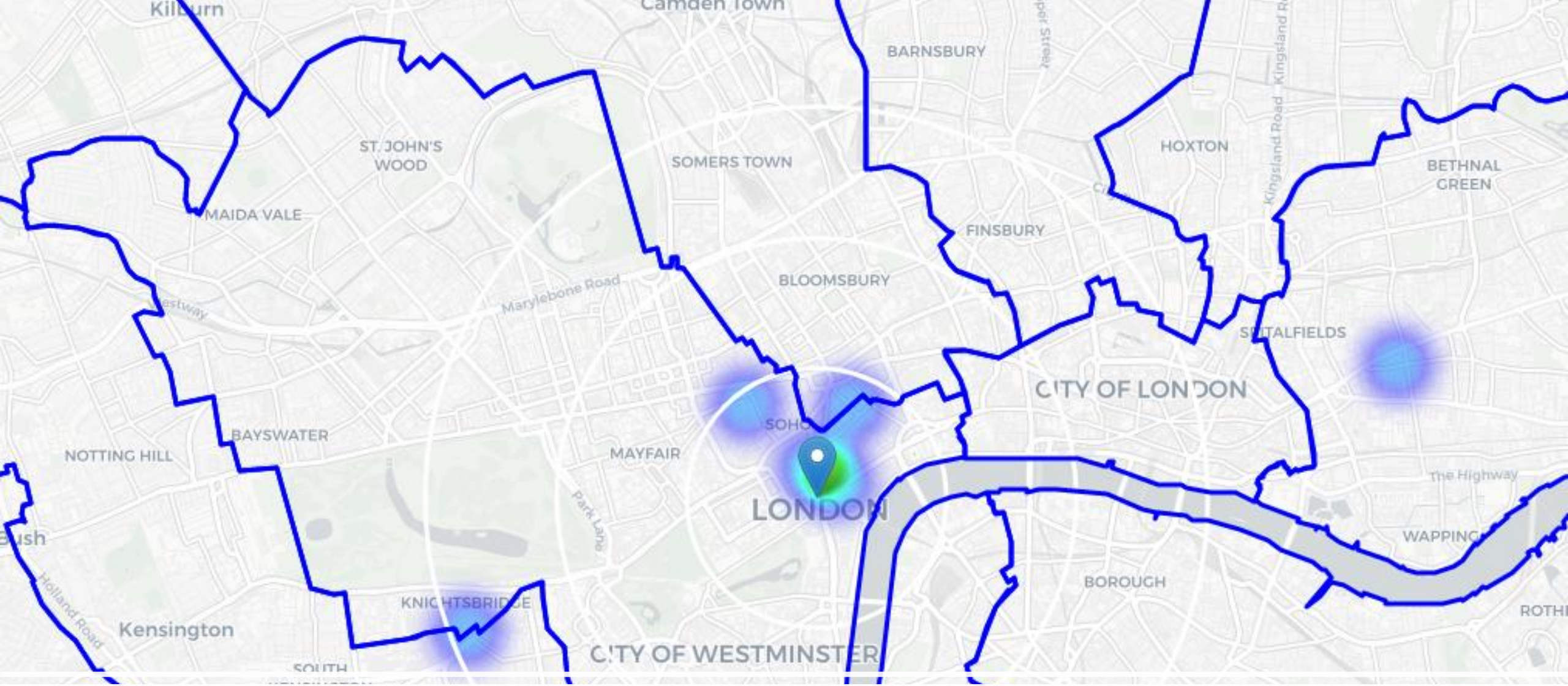
- In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **no more than two restaurants in radius of 250 meters**, and we want locations **without Indian 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.

Analysis

- Average number of restaurants in every area with radius=300m: 0.9725274725274725
- **On average Indian restaurant can be found within ~2750m** from every area centre candidate



heatmap / density of restaurants



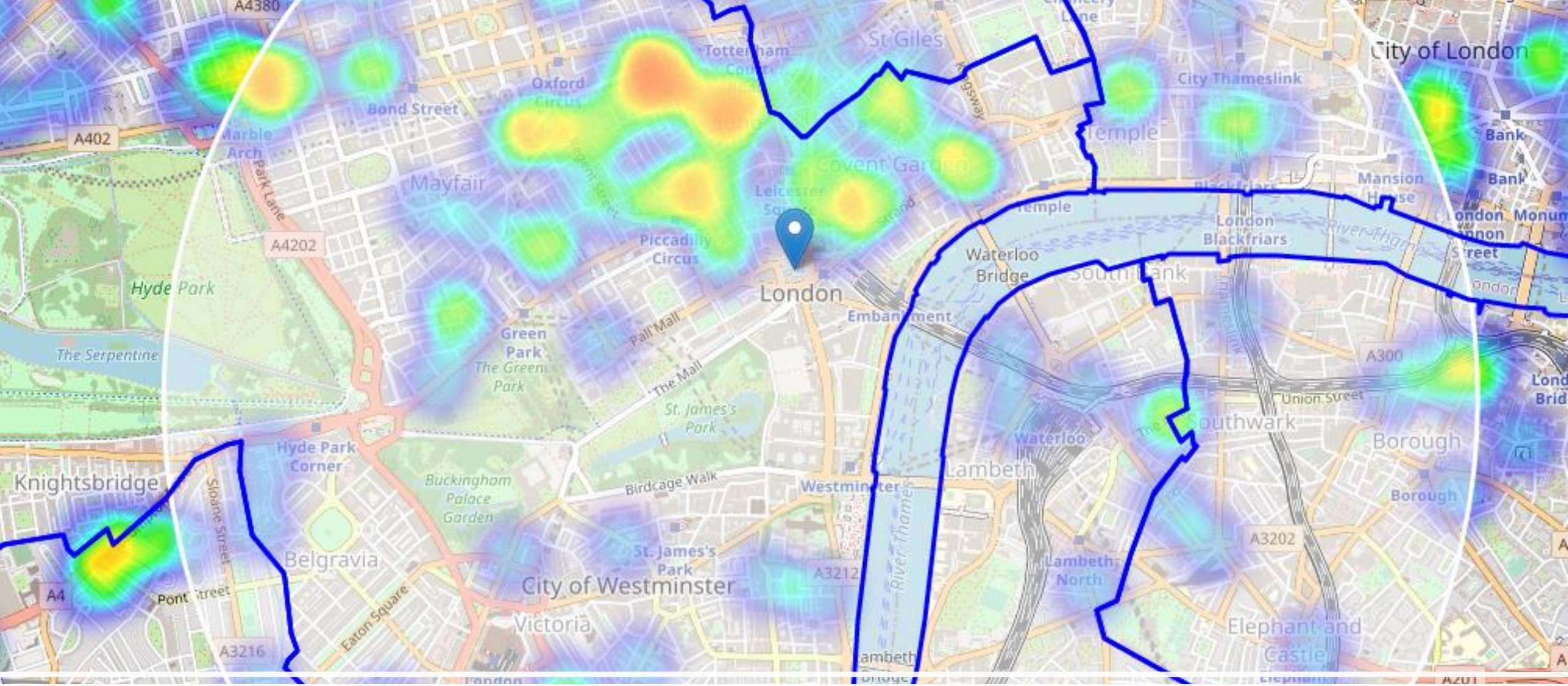
Heatmap/density of Indian restaurants only.

Inferences:

- Looks like a few pockets of low restaurant density closest to city centre can be found **south, west, north and east from Trafalgar Square**.
- This map is not so 'hot' (Indian restaurants represent a subset of ~2% of all restaurants in London) but it also indicates higher density of existing Indian restaurants directly north and north-east from Trafalgar Square, with closest pockets of **low Indian restaurant density positioned west, south-west, south and south-east from city centre**.
- Based on this we will now focus our analysis on areas *a slightly south from London centre* - we will move the centre of our area of interest and reduce its size to have a radius of **2.5km**. This places our location candidates mostly in boroughs **London Borough of Westminster, London Borough of Lambeth, London Borough of Southwark and London Borough of Camden**

London Borough of Westminster

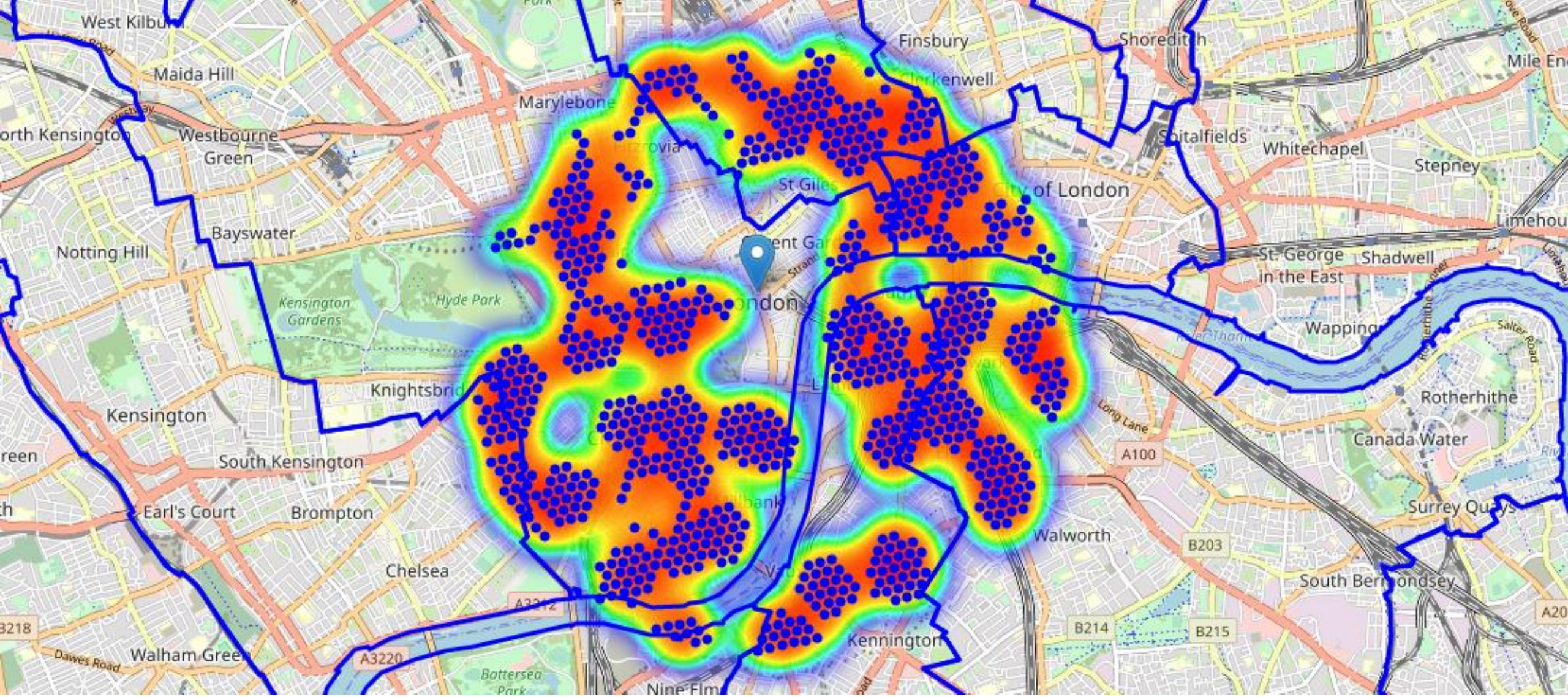
- "Westminster is London's most famous borough and forms the beating heart of our capital. It is known and recognised throughout the world as the location for some of the city's most spectacular sites; 'Big Ben' and the Houses of Parliament, Buckingham Palace, Trafalgar Square and Westminster Abbey. It is where the UK's central government is based, it is home to the Queen and other members of the Royal Family, and is the centre of ceremony and pageantry."
- "Famous for being the centre of art and culture, leisure and entertainment, shopping and style, it is where many important institutions have their headquarters and is probably the most exciting area of London."
- "It is one of London's 33 boroughs and is bordered by the boroughs of Camden to the north, the City of London to the east, by Kensington & Chelsea and Brent to west, and the river Thames to the south. It covers just over 8 square miles of central London and was created in 1965, with the unification of three former Metropolitan boroughs, Paddington, Marylebone and Westminster. Around 240,000 people call Westminster home, and 3,500 new residents are born here every year. Every day over a million people visit, whether to work, study, or to see the sights." (https://westminsterguides.org.uk/)



Low-restaurant-count parts of City of Westminster closest to Trafalgar Square

Denser grid of location candidates restricted to our new region of interest

- 2261 candidate neighborhood centres generated.
- Locations with no more than two restaurants nearby: 888
- Locations with no Indian restaurants within 400m: 2081
- Locations with both conditions met: 859

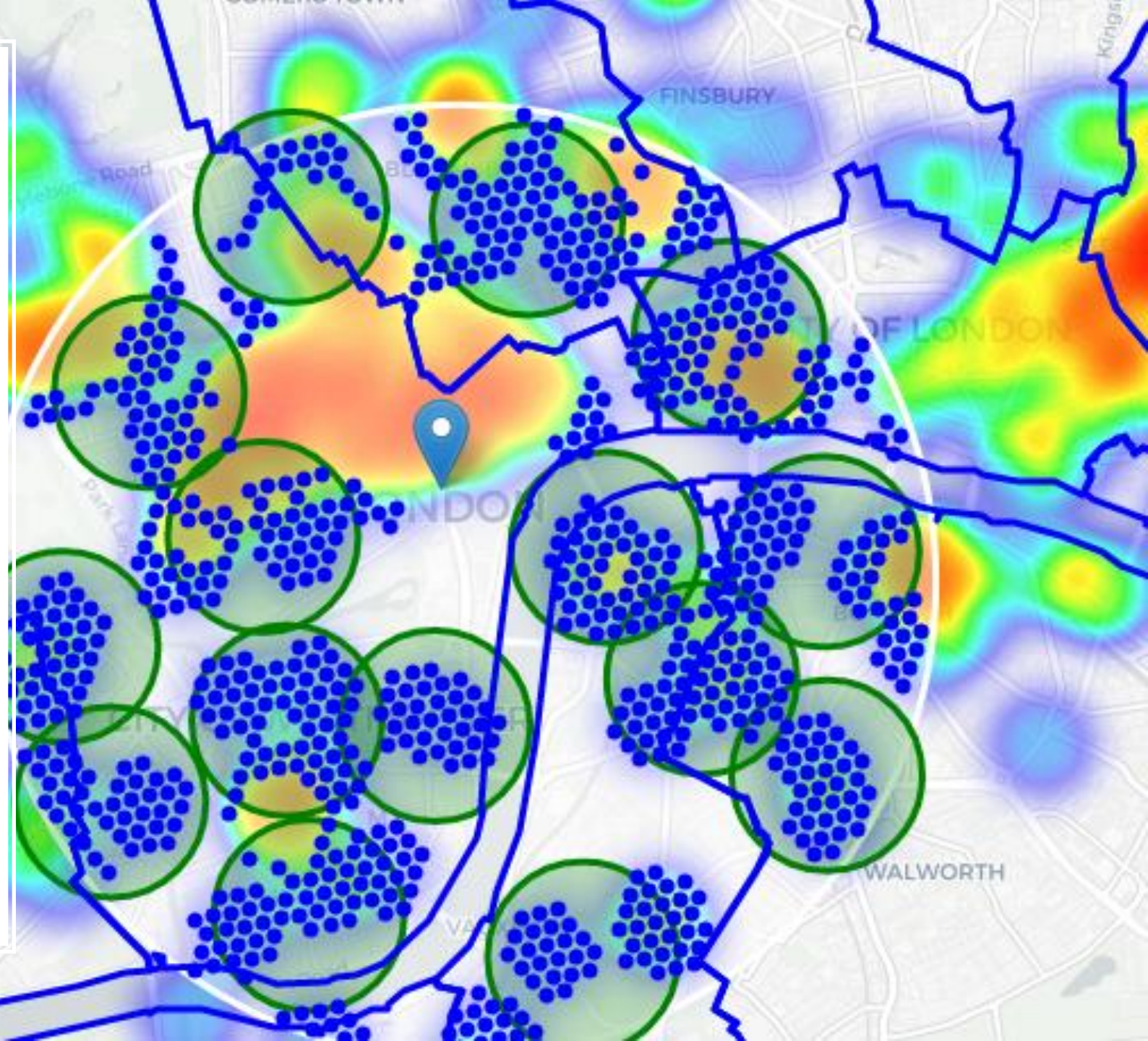


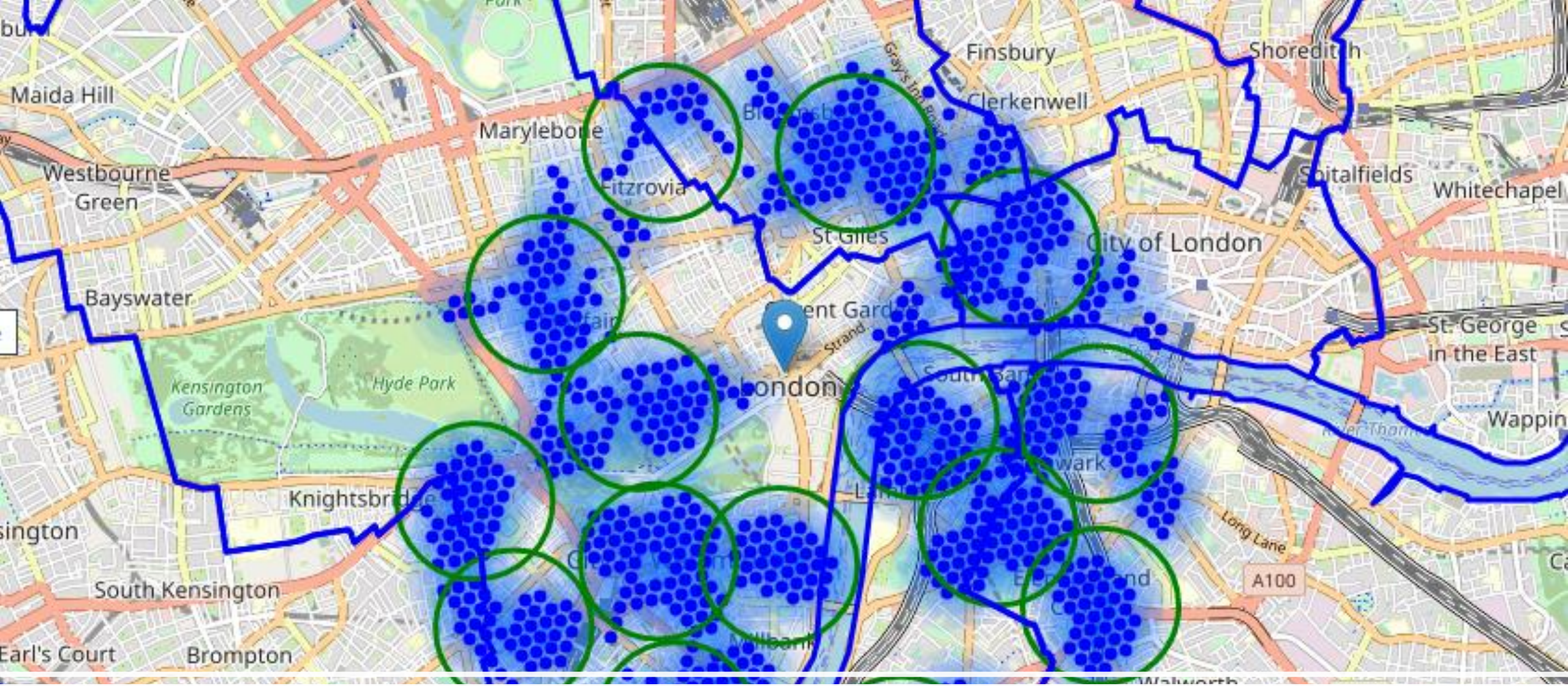
Good locations in a form of heatmap

Cluster the locations to create **centres of zones containing good locations**.

Those zones, their centres and addresses will be the result of our analysis.

- Our clusters represent groupings of most of the candidate locations and cluster centres are placed nicely in the middle of the zones 'rich' with location candidates.
- Addresses of those cluster centres will be a good starting point for exploring the neighbourhoods to find the best possible location based on neighbourhood specifics.

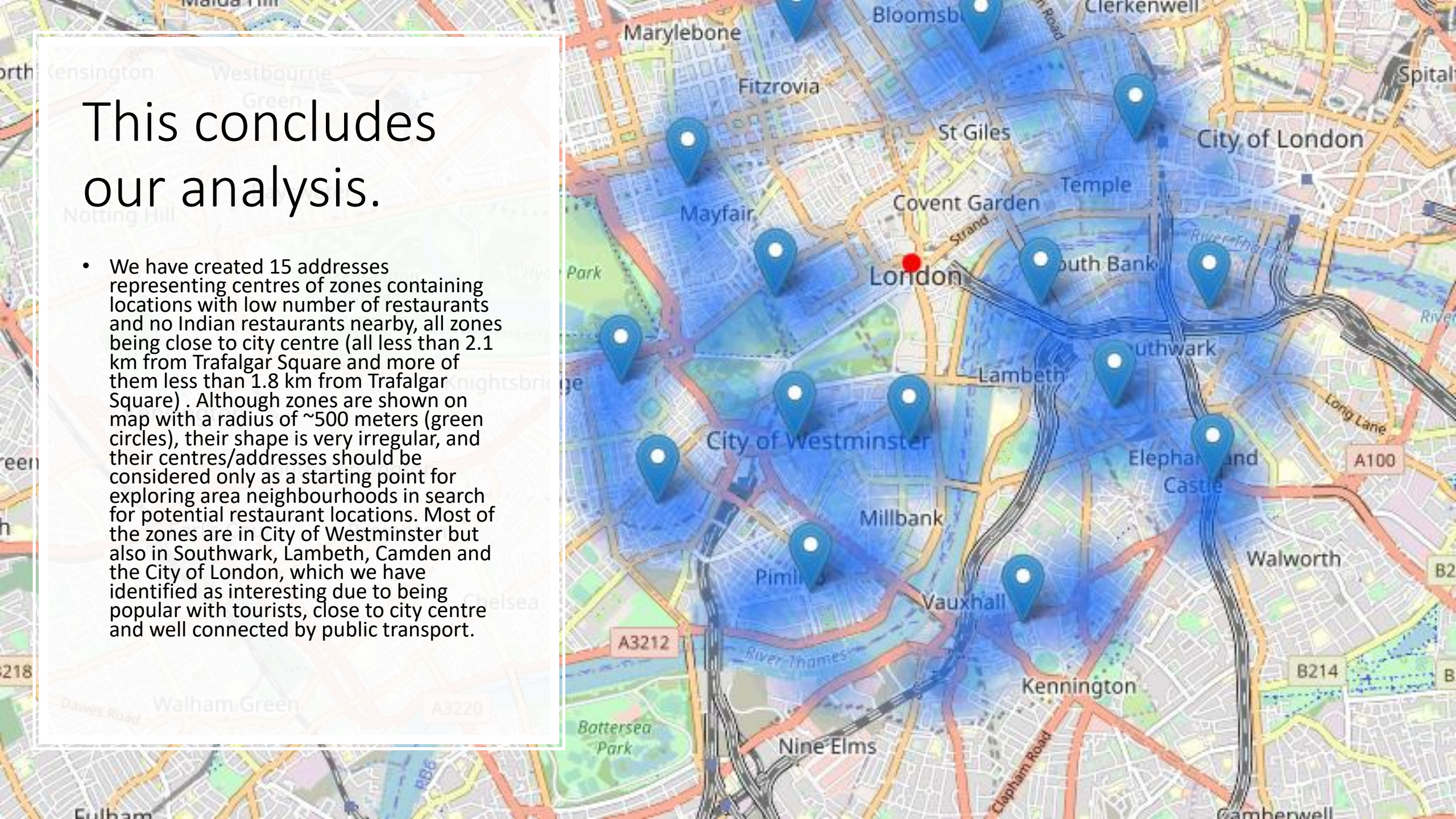




Visualise the zones on a city map without heatmap, using shaded areas to indicate our clusters:

This concludes our analysis.

- We have created 15 addresses representing centres of zones containing locations with low number of restaurants and no Indian restaurants nearby, all zones being close to city centre (all less than 2.1 km from Trafalgar Square and more of them less than 1.8 km from Trafalgar Square) . Although zones are shown on map with a radius of ~500 meters (green circles), their shape is very irregular, and their centres/addresses should be considered only as a starting point for exploring area neighbourhoods in search for potential restaurant locations. Most of the zones are in City of Westminster but also in Southwark, Lambeth, Camden and the City of London, which we have identified as interesting due to being popular with tourists, close to city centre and well connected by public transport.



Results and Discussion

- Our analysis shows that although there is a great number of restaurants in London (~378 in our initial area of interest which was 12x12km around Trafalgar Square), there are pockets of low restaurant density close to city centre. Highest concentration of restaurants was detected north and north-east and north-west from Trafalgar Square, so we focused our attention to areas south, corresponding to borough of Westminster and north-western corner of Southwark, northern corner of Lambeth, southern corner of Camden and east corner of the City of London. Our attention was focused on Westminster which offer a combination of popularity among tourists, closeness to city centre, strong socio-economic dynamics and several pockets of low restaurant density.
- After directing our attention to this more narrow area of interest (covering approx. 5x5km south from Trafalgar Square) we first created a dense grid of location candidates (spaced 100m apart); those locations were then filtered so that areas with no restaurants and those with more than two restaurants in radius of 250m also those with an Indian restaurant closer than 400m were removed.

Results and Discussion

- 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.
- 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 Indian restaurants particularly. This, of course, does not imply that those zones are optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to London centre but not crowded with existing restaurants (particularly Indian) - 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 low competition, but also other factors considered, 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 Indian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Indian restaurant. By calculating restaurant density distribution from Foursquare data, we have first identified general boroughs that justify further analysis (Westminster), 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.