

EXPLORING LOCATIONS TO OPEN AN ITALIAN RESTAURANT IN LONDON

IBM Data Science Capstone Project

Pavel Gotovski

Email: pavelgto@gmail.com
Github: pavelg0t

Date:
2020-04-02

Table of contents

1. Introduction
2. Data
 - 2.1. Feature selection
 - 2.2. Data sources
 - 2.3. Data acquisition
 - 2.3.1. London business districts
 - 2.3.2. Metro stations in the area
 - 2.3.3. Food venues nearby
3. Methodology
4. Analysis
 - 4.1. Distribution of food venues per area
 - 4.2. Potential candidates
 - 4.3. Choosing final candidate
5. Results and Discussion
6. Conclusion

Introduction

Background

London is the capital and largest city of England and the third-most populous city in Europe. It is considered to be one of the world's most visited, most expensive and most influential cities. It is known for its rainy climate, ethnic diversity, Big Ben, the financial industry, and its sprawling subway system.

London is also one of the most amazing places to eat, thanks to an incredible variety of international cuisines and some of the most talented chefs in the world. London's thriving economy, multicultural demographics and access to ingredients make it an ideal place for restaurants to flourish.

Business problem

The objective of this project is to find the most suitable location(s) to open a new Italian restaurant in London.

Target audience

Types of clients or a group of people that would be interested in this project are:

- Business personnel who wants to invest or open an Italian restaurant in London. This analysis will be a comprehensive guide to start or expand restaurants targeting the large pool of office workers in major business districts of London city during lunch hours.

Data

Feature selection

Based on definition of the problem, factors that are considered influential:

- Regular flow of customers daily.
- As few rivals as possible.
- Good transportation system/connection to other neighborhoods of the city.

Decision was made to look for potential places in the major business districts of London city. It should be within a walking distance to a nearby metro stations, so areas within 500 meters from every metro station are considered. Next, locations of already existing restaurants in the vicinity are gathered. The most prominent candidate area is chosen based on the food venue density and proximity to the center of the business district. Potential places for a new Italian restaurant within the chosen area are selected based on these criteria: not closer to other Italian restaurants within 200 meters and other food venues within 100 meters.

Data sources

To extract/generate required information the following data sources were used:

- [Research article on major business districts in London](#).
- Locations of metro stations and different cuisine restaurants gathered with the help of [Foursquare API](#).
- Reverse geocoding (coordinates transformation to address) using [LocationIQ API](#).

Data acquisition

London business districts

Although there are [five major business districts in London](#), only the biggest three were chosen: **West End**, **The City**, and **Canary Wharf**. Their areas are shown in Fig. 1

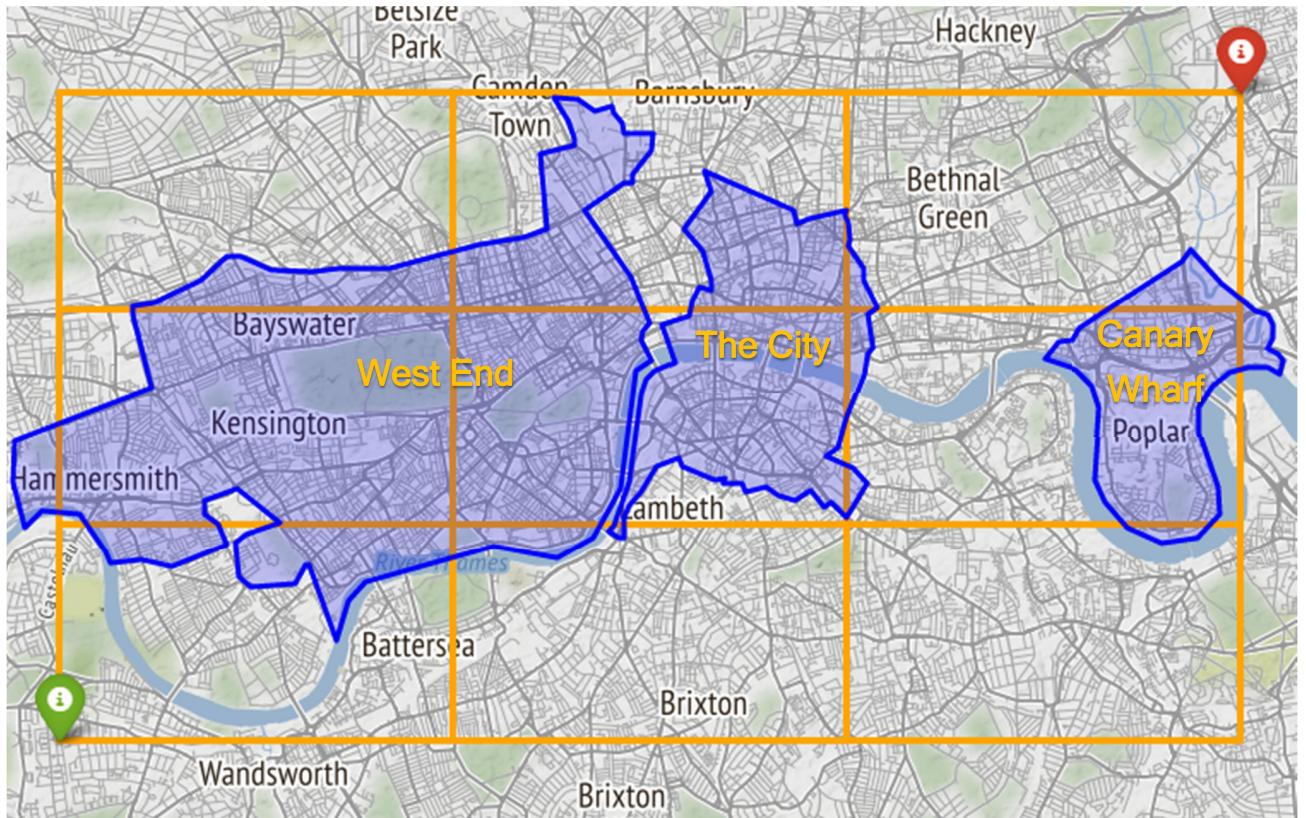


Figure 1. Three major business districts in London. Their combined area is approximated as a rectangle.

To simplify the search for any metro station in the business areas, combined territories of all three districts were approximated as a rectangle. Coordinates of the rectangle's south-west (green marker) and north-east (red marker) corners are:

sw: (51.461,-0.236),

ne: (51.542,0.001).

Metro stations in the area

Foursquare API was used to gather locations of all metro stations in the rectangular area. In the Foursquare's Place API all venues are categorized into [10 parent categories](#) (each having its own subcategories). To make an API call, venues under category name 'Metro Station' and categoryID code '4bf58dd8d48988d1fd931735' were chosen. Total of 116 venues were found and 107 of them were underground stations. List of the few metro stations found is shown in the Tab. 1.

Table 1 An example of few metro stations found in the rectangular area.

	name	latitude	longitude
0	Baker Street London Underground Station	51.5230	-0.1572
1	Edgware Road London Underground Station	51.5202	-0.1670
2	Paddington London Underground Station	51.5165	-0.1767
...
104	Clapham North London Underground Station	51.4650	-0.1295
105	Nine Elms London Underground Station	51.4803	-0.1293
106	Battersea Power Station London Underground Station	51.4799	-0.1420

Locations of all metro stations are shown in Fig. 2

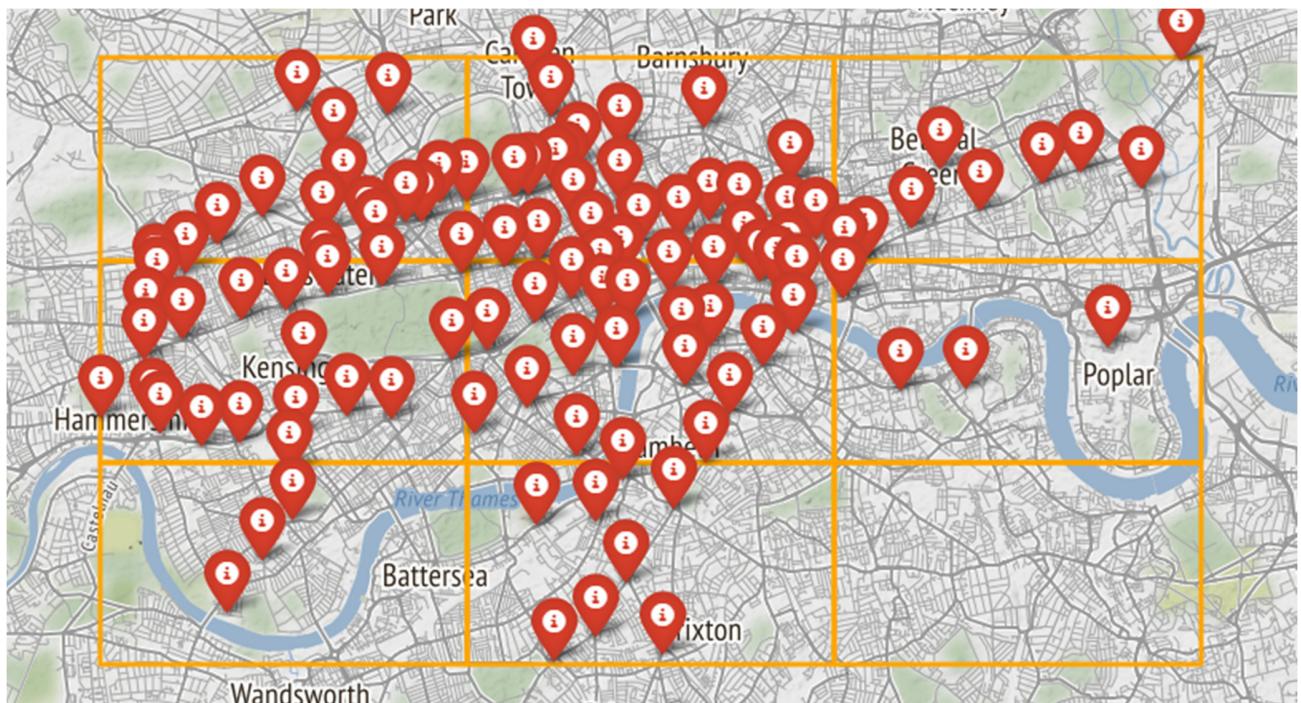


Figure 2. Locations (red marker) of all metro stations found in the rectangular area.

Food venues nearby

As in the section before, Foursquare's Place API was used to find locations of all food venues (in particular Italian restaurants) within radius of 500 meters from every metro station found previously. Venues under category name 'Food venues' (categoryID '4d4b7105d754a06374d81259') were specified in the API calls. Total of 15084 venues were found (spread through 246 different subcategories) and 831 of them were categorized as 'Italian restaurant'.

An example of food venues' locations (green marker) around the Waterloo London Underground Station is shown in Fig. 3

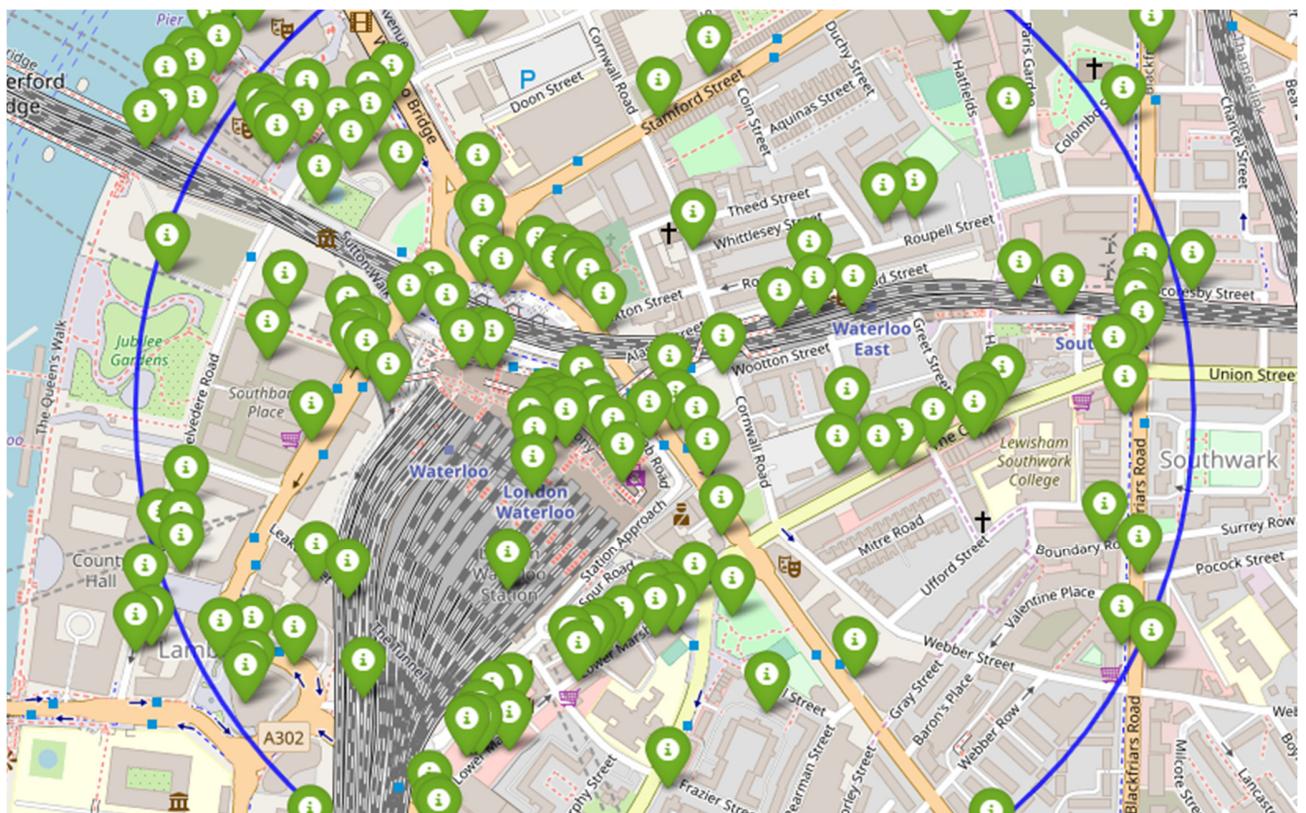


Figure 3. Locations of all food venues nearby the Waterloo London underground station.

Methodology

In the data acquisition section, required data was gathered: location and type (category) of any food venue (including Italian restaurant) within ~500 meters from every metro station in our defined rectangular area.

In the analysis section, calculations and exploration of 'food venue density' (and also 'Italian restaurant density') across different areas are done. Based on these density measures (low number of restaurants nearby) the potential candidates' list is created.

After final candidate is chosen, further analysis is made focusing on most promising places for a new Italian restaurant. These places must satisfy basic requirements defined earlier: locations with no food venues within a radius of 100 meters, and no other Italian restaurant within a radius of 200 meters.

A map of all such locations is presented and their general clusters are calculated. Centers of these clusters will be used as a starting point for final 'street-level' exploration and search for optimal venue location.

Analysis

Distribution of food venues per area

Distributions of Italian restaurants and food venues around metro stations in the rectangular area are shown in Fig. 4

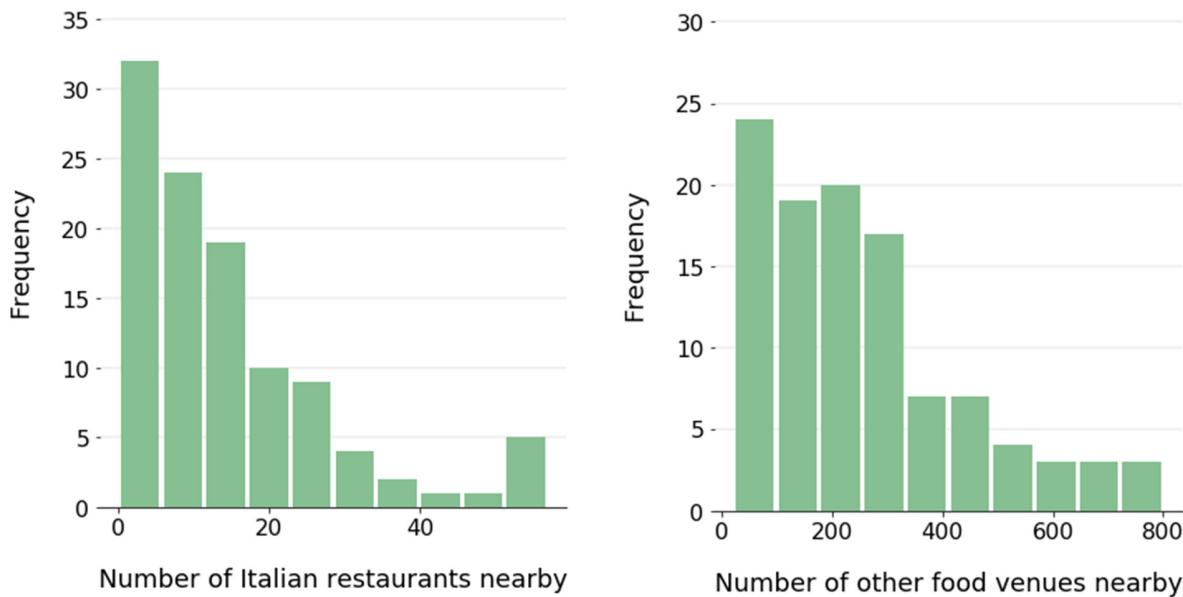


Figure 4. Distribution of Italian restaurants (left) and other food venues (right) per area.

Basic food venue distribution statistics are shown in Tab. 2

Table 2 Basic Italian restaurant and other food venue distribution statistics.

	Italian restaurants	Other food venues
mean	14.7	253.5
std	13.4	182.6
min	0.0	20.0
25%	5.0	113.0
50%	11.0	214.0
75%	20.5	334.0
max	57.0	801.0

*Important thing to notice: The total number of food venues displayed in Fig. 4 is greater than the actual 15084 (there are multiple duplicates of same venues in the data set due to the fact that areas around few of the stations are overlapping).

Potential area candidates

List of potential candidate areas is created based on these criteria:

- The number of Italian restaurants is equal or less than 6 (around first quartile).
- The number of other food venues is equal or less than 214 (around median).

Total of 32 areas meet the criteria. The list is displayed in Tab. 3

Table 3 List of potential candidate areas.

	Station name	Italian restaurants	Food venues	latitude	longitude	cluster
0	Barons Court	0	65	51.4904	-0.2141	0
1	Battersea Power Station	3	43	51.4799	-0.1420	2
2	Bermondsey	3	37	51.4978	-0.0637	1
3	Bethnal Green	3	106	51.5274	-0.0554	1
4	Bow Road	0	23	51.5270	-0.0252	1
5	Bromley-By-Bow	1	20	51.5247	-0.0122	1
6	Canada Water	0	59	51.4980	-0.0497	1
7	Clapham Common	5	118	51.4617	-0.1382	2
8	Clapham North	4	91	51.4650	-0.1295	2
9	Fulham Broadway	5	158	51.4805	-0.1947	0
10	Hammersmith	5	187	51.4938	-0.2250	0
11	Hammersmith	6	184	51.4920	-0.2231	0
12	Holland Park	4	47	51.5073	-0.2055	0
13	Kennington	3	62	51.4883	-0.1058	2
14	Kilburn Park	4	41	51.5350	-0.1937	0
15	Latimer Road	4	41	51.5135	-0.2177	0
16	Maida Vale	0	41	51.5299	-0.1858	0
17	Mile End	2	78	51.5253	-0.0335	1
18	Nine Elms	2	73	51.4803	-0.1293	2
19	Oval	2	75	51.4821	-0.1126	2
20	Pimlico	6	114	51.4892	-0.1335	2
21	St. John's Wood	2	42	51.5346	-0.1741	0
22	Stepney Green	3	82	51.5219	-0.0467	1
23	Stockwell	0	37	51.4723	-0.1228	2
24	Vauxhall	3	74	51.4859	-0.1237	2
25	Warwick Avenue	3	92	51.5233	-0.1837	0
26	West Brompton	3	88	51.4870	-0.1953	0
27	West Kensington	3	83	51.4907	-0.2060	0
28	Westminster	4	112	51.5009	-0.1248	2
29	White City	2	123	51.5116	-0.2241	0
30	Whitechapel	2	132	51.5195	-0.0615	1
31	Wood Lane	5	195	51.5101	-0.2238	0

The potential areas are further clustered based on their geographical position using k-Means clustering algorithm. Total of 3 clusters where identified using elbow method. These clusters are shown in Fig. 5

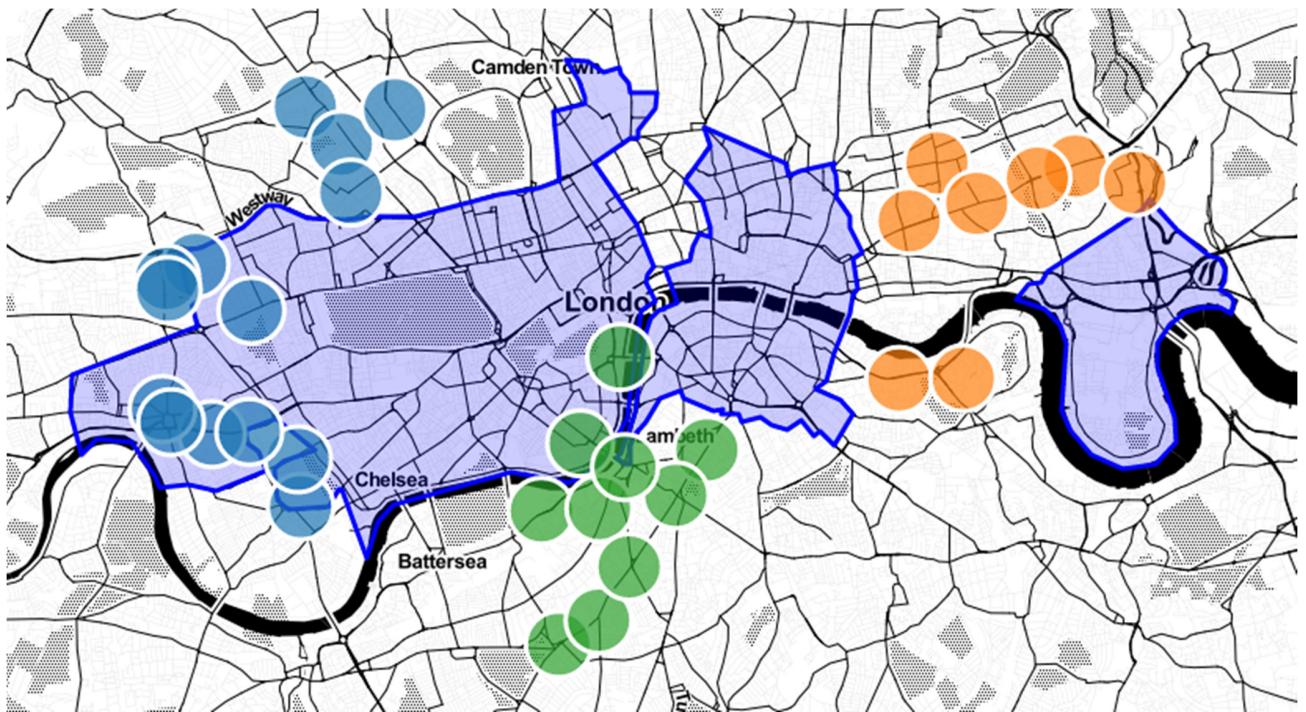


Figure 5. Clusters of potential candidate areas.

Choosing Final candidate

As shown in Fig.5 areas within the blue cluster overlap with the business districts' areas the most and seem to be the most promising at first. There are only two candidates from the green cluster and none from the orange.

Further analysis of the blue cluster's candidates reveal that all of them are situated in the low business density area called 'Fringe' in the [Savills reaserch](#).

On the other hand, the two candidates from the green cluster are situated on the edge of the densest part called 'Core' so they were chosen as the most suitable candidates. This areas are around Pimlico and Westminster underground stations.

Finally, area surrounding the Pimlico underground station was chosen as the most promising based solely on the assumption of high rental cost in the neighboring area around Westminster station (buildings like Big Ben and UK Parliament attract lots of tourists and high prices).

Potential locations in the Pimlico/Millbank area

The Pimlico/Millbank area was analyzed further by creating a dense grid of location candidates (spaced ~50m apart) and finding the places located at least 200m away from any other Italian restaurant and 100 meters from any other food venue. Density of food venues and potential places for a new Italian restaurant are shown in Fig. 6

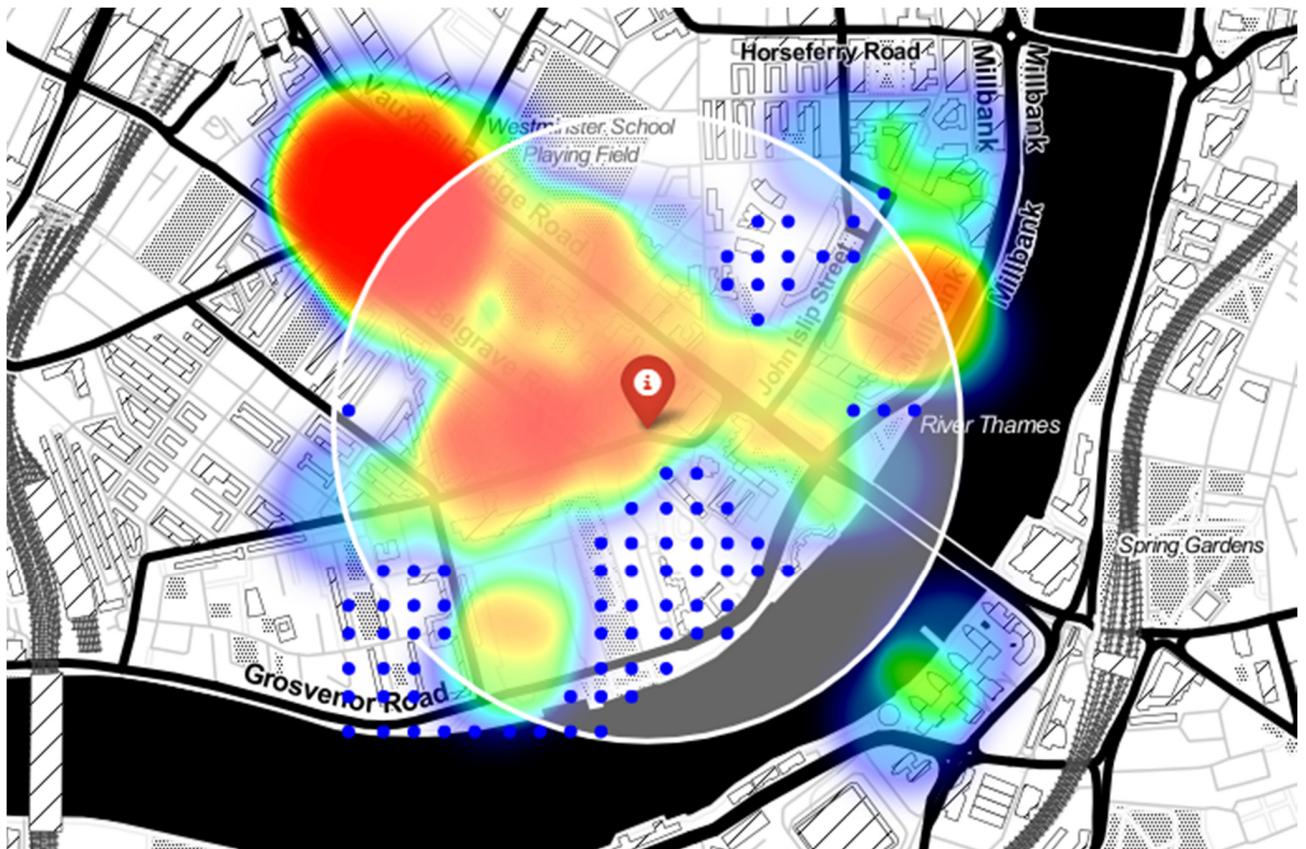


Figure 6. Map of food venue density (rainbow) and potential locations for a new Italian restaurant (blue dots) around the Pimlico underground station (red marker). White line indicates a 500m radius circle area.

Potential locations were clustered into zones using k-Means clustering method used before. Total of 4 zones with a large number of potential new restaurant places were identified. These clusters are shown in Fig. 7.



Figure 7. Zones (blue marker) with a large number of potential new restaurant places (blue dot) nearby the Pimlico underground station (red marker).

Addresses of these zones were found using reverse geocoding and are shown in Tab. 4

Table 4 Addresses of final 4 zones in the Pimlico/Millbank area.

	address	latitude	longitude	distance to station (m)
0	Balvaird Place, Millbank, London	51.4873	-0.1323	220
1	Churchill Gardens Road, Pimlico, London	51.4860	-0.1393	530
2	Erasmus Street, Millbank, London	51.4912	-0.1299	340
3	St George's Square, Pimlico, London	51.4859	-0.1346	370

Results and Discussion

Our analysis shows that there are 106 underground metro stations in the rectangular area of 16x9 km representing major business districts in London. On average there are around 14 Italian restaurants and 253 other food venues within the radius of ~500 meters from the stations. We focused our attention on areas with the least number of food venues and closest to the centers of the business districts.

At first glance, the Hammersmith, West Kensington and Walham Green areas (also known as a part of the 'Fringe' area in the savills research) seem to be the best candidates but when the amount of office sq. ft. by area is taken into account, the two areas around Pimlico and Westminster underground stations are chosen as the most promising. The Pimlico area became the final candidate because of the likely hudge rental cost in the Westminister area.

The Pimlico area was analyzed further by creating a dense grid of location candidates (spaced ~50m apart) finding the places located at least 200m away from any other Italian restaurant and 100 meters from any other food venue.

Location candidates were clustered to create general zones of interest. Addresses of these zones were found using reverse geocoding.

The result of this analysis is 4 zones with a large number of potential new restaurant places. Of course, these zones are not optimal locations for a new restaurant. There may be very good reasons for a small number of food venues in these locations (e.x. locations pointing to places like a park, a river or a road where it is impossible to build or rent. Negative neighbourhood characteristics like pollution, noise and etc.) which would make them unsuitable for a new restaurant. Recommended zones should be considered only as a starting point for a more detailed 'street-level' analysis.

Conclusion

The purpose of this project was to identify areas close to a metro station with a low number of food venues (in particular Italian restaurants) to aid future investors in finding the potential location for a new Italian restaurant. By aiming at office workers as the future daily clients, only the metro stations situated in London's major business districts were taken into consideration.

By calculating food venue density distribution from Foursquare API data the most suitable candidate (Pimlico/Millbank area) was identified, and a further analysis conducted. An extensive collection of locations that satisfy some basic requirements regarding existing nearby restaurants was generated. These locations were clustered into 4 zones with their unique addresses.

A final decision on optimal restaurant location will be made by future investors based on specific characteristics of locations in every recommended zone, taking into consideration additional factors like proximity to a business center and major roads, rental prices, and other economic variables.