

A new Italian restaurant in London

Using the Foursquare location data to predict
the best area of London to open an Italian restaurant

Filippo M.

January 16, 2021

1 Introduction

1.1 Background

London is one of the most multicultural cities in the world, with 36.7% of the population being foreign-born [3] and a multitude of tourists visiting the city every year. According to the *2011 United Kingdom Census* [3], Italian citizens represent one of the major immigration groups in London. The cultural diversity has an influence on the vast choice of different restaurants in the city, with Italian cuisine certainly being extremely popular.

1.2 Problem

The popularity of Italian restaurants in London might be seen as an obstacle for entrepreneurs that are interested in opening such type of restaurant. Popularity might translate into strong competition, and therefore into difficulties in starting a successful business. Data can help understanding which area of the city is best suited to welcome a new Italian restaurant, as it can help identifying where the competition is weaker and point new entrepreneurs in the right direction.

1.3 Interest

The target audience of this report is any restaurant entrepreneur with an interest in starting an Italian restaurant business in the London area. Clearly, depending on the available budget level more considerations could be added on top of those presented in this report.

2 Data

2.1 Data Sources

The data processed in this report have been taken from two main sources. The first source is a Wikipedia page which lists all the Boroughs of London, with their geographical coordinates and their categorization as part of either the Inner London or the Outer London [1]. Secondly, this report uses Foursquare location data. More specifically, it takes advantage of the Foursquare *explore endpoint* [4], which returns a list of popular locations for a given set of latitude and longitude coordinates. The information available in this list include their category, which is paramount for the scope of this analysis as it allows to enumerate restaurants (including Italian restaurants) in each given Borough.

2.2 Assumptions

Before proceeding to the analysis of the data, some considerations and assumptions need to be discussed. It has been mentioned in 2.1, the distinction between Inner London and Outer London. The first are the Borough of London that are located in the most central part of the city [2]. They attract tourists from all over the globe and are the center of the financial activities of the city, with several headquarters of multinational groups located within their boundaries. Therefore, large numbers of people commute everyday towards these central locations, opening possibilities for restaurant businesses to succeed. The Outer London does not have the same demographic characteristics, with a density of population that is less than half of the most central areas [2]. Generally, the Outer London Boroughs do not see a comparable flux of people entering on a daily basis. For these reasons, this report will try and make a distinction between the two areas, reporting results and comparisons both with a generic approach (comparing all Borough regardless of their Inner/Outer status) and also separating the two groups.

Some Boroughs represent an anomaly in the categorization between Inner and Outer London, as their status varies between the statutory definition and the statistical definition. This report complies with the statistical definition, therefore Haringey and Newham are considered Inner London, while Greenwich is Outer London [1, 2]. Furthermore, City of London is not officially categorized as a Borough, but it has been included in this analysis as part of the Inner London in order to provide a complete view of all areas of the city.

Another important assumption regards the grouping and categorization of restaurants. As previously mentioned, Foursquare will provide the type of location for each popular place. In order to perform an analysis on the popularity of restaurants in each area, different types of categories that are not strictly defined as *restaurants* have been counted as restaurants, since they represent a form of competition for a potential new business. This includes, for examples, locations that are defined as *Burrito Place*, *Food Truck*, *Noodle House* and *Food Court*. Finally, the specific popularity of Italian restaurants needs to be evaluated. A decision has been taken to include all locations that are marked as *Pizza Place* into the group of *Italian Restaurants*. Pizzerias are normally part of the Italian cuisine tradition, and therefore to the eye of a customer they represent an alternative to an Italian restaurant within the context of Italian cuisine. This is a very important assumption, as the popularity and the presence of Italian restaurants in each area is an important parameter of the analysis.

2.3 Data Pre-processing

Before being discussed, data have been selected and filtered. Specifically, the data from the Wikipedia table [1] that have been used in this report are the names, coordinates and Inner/Outer status of each Borough. Other pieces of information such as population, area, headquarters location and political orientation have been omitted as they did not serve the purpose of the report.

In using the Foursquare explore endpoint, a radius can be given as an input to the function, in order to decide the area of influence of the Foursquare investigation. Due to the vastness of most of the Outer London Boroughs (if compared to the more compact Inner London Boroughs) a larger radius has been selected to query popular locations in the outer areas with respect to the central ones.

3 Methodology

3.1 Parameters

Firstly, it is important to discuss the two main parameters that have been used to assess the feasibility of a new Italian restaurant in each Borough. Both these parameters rely on the idea that it is not easy to start a new business where there is a strong competition. Therefore, a low number of competitors among the 100 most popular venues is to be considered a positive aspect. The first parameter is the frequency with which restaurants occur amongst the top venues for a given Borough. This parameter has been named *RF* (*Restaurant Frequency*), as it includes all types of restaurants in the calculation. The second parameter

aims at quantifying the competition with other restaurants specialized in Italian cuisine. It has been named *IRF* (*Italian Restaurant Frequency*), and it consists in the frequency with which Italian restaurants occur amongst the popular locations. In formula, for each i^{th} Borough these two parameters read as follow:

$$RF_i = \frac{\text{Number of restaurants among the top venues in the } i^{\text{th}} \text{ Borough}}{\text{Ovreal number of top venues found for the } i^{\text{th}} \text{ Borough}}$$

$$IRF_i = \frac{\text{Number of Italian restaurants among the top venues in the } i^{\text{th}} \text{ Borough}}{\text{Ovreal number of top venues found for the } i^{\text{th}} \text{ Borough}}$$

The lower these two parameters are, the weaker is the competition, as the number of similar businesses among popular location is lower.

Practically, these parameters are obtained by applying hot one encoding to the database, grouping and averaging for each Borough. This results in the frequency with which every category appears, including the categroy *Italian Restaurant*. Finally, by summing the frequencies of the categories that have the word *Restaurant*, the *RF* is found.

3.2 Exploratory Analysis

The first opportunity to observe the data is given by the map with all the Borough locations 2. The Python code provided by the author has an interactive map, with the possibility to visualize the names of the Boroughs by clicking on it. Secondly, all the *RF* and the *IRF* values are reported in the full results table 1. This table includes all the most important information: Inner/Outer status, *IRF*, *RF* and Borough name, but it is not the best data visualization tool. This is because data are sorted alphabetically, and they are not helping the reader understand the message and capture the most interesting aspects of this data. However, the table 1 in combination with the map 2, can help with a good preliminary observation.

This data were observed and studied before proceeding, in order to understand if relevant information could be extracted by it. Furthermore, different radii of Foursquare investigation were tested, before selecting the one that gave the highest amount of locations. It is finally important to mention that no clustering or categorization method was used in this report, as it would not serve the specific scope of the analysis.

	Borough	All Restaurants	Italian Restaurant	Inner	Latitude	Longitude
0	Barking and Dagenham	0.149254	0.044776	Outer London	51.5607	0.1557
1	Barnet	0.355556	0.077778	Outer London	51.6252	-0.1517
2	Bexley	0.200000	0.046154	Outer London	51.4549	0.1505
3	Brent	0.320000	0.050000	Outer London	51.5588	-0.2817
4	Bromley	0.348837	0.069767	Outer London	51.4039	0.0198
5	Camden	0.290000	0.050000	Inner London	51.5290	-0.1255
6	City of London	0.290000	0.050000	Inner London	51.5155	-0.0922
7	Croydon	0.362500	0.025000	Outer London	51.3714	-0.0977
8	Ealing	0.410000	0.090000	Outer London	51.5130	-0.3089
9	Enfield	0.225806	0.048387	Outer London	51.6538	-0.0799
10	Greenwich	0.151515	0.045455	Outer London	51.4892	0.0648
11	Hackney	0.220000	0.030000	Inner London	51.5450	-0.0553
12	Hammersmith and Fulham	0.390000	0.080000	Inner London	51.4927	-0.2339
13	Haringey	0.320000	0.020000	Inner London	51.6000	-0.1119
14	Harrow	0.371134	0.051546	Outer London	51.5898	-0.3346
15	Havering	0.183333	0.050000	Outer London	51.5812	0.1837
16	Hillingdon	0.232143	0.017857	Outer London	51.5441	-0.4760
17	Hounslow	0.315217	0.021739	Outer London	51.4746	-0.3680
18	Islington	0.310000	0.040000	Inner London	51.5416	-0.1022
19	Kensington and Chelsea	0.370000	0.060000	Inner London	51.5020	-0.1947
20	Kingston upon Thames	0.255556	0.022222	Outer London	51.4085	-0.3064
21	Lambeth	0.402299	0.034483	Inner London	51.4607	-0.1163
22	Lewisham	0.220000	0.080000	Inner London	51.4452	-0.0209
23	Merton	0.380282	0.042254	Outer London	51.4014	-0.1958
24	Newham	0.163636	0.018182	Inner London	51.5077	0.0469
25	Redbridge	0.293478	0.054348	Outer London	51.5590	0.0741
26	Richmond upon Thames	0.230000	0.040000	Outer London	51.4479	-0.3260
27	Southwark	0.300000	0.060000	Inner London	51.5035	-0.0804
28	Sutton	0.196970	0.090909	Outer London	51.3618	-0.1945
29	Tower Hamlets	0.310000	0.050000	Inner London	51.5099	-0.0059
30	Waltham Forest	0.250000	0.040000	Outer London	51.5908	-0.0134
31	Wandsworth	0.204545	0.034091	Inner London	51.4567	-0.1910
32	Westminster	0.250000	0.030000	Inner London	51.4973	-0.1372

Figure 1: Full Table. *Italian Restaurant* and *All Restaurants* are the values of *IRF* and *RF*

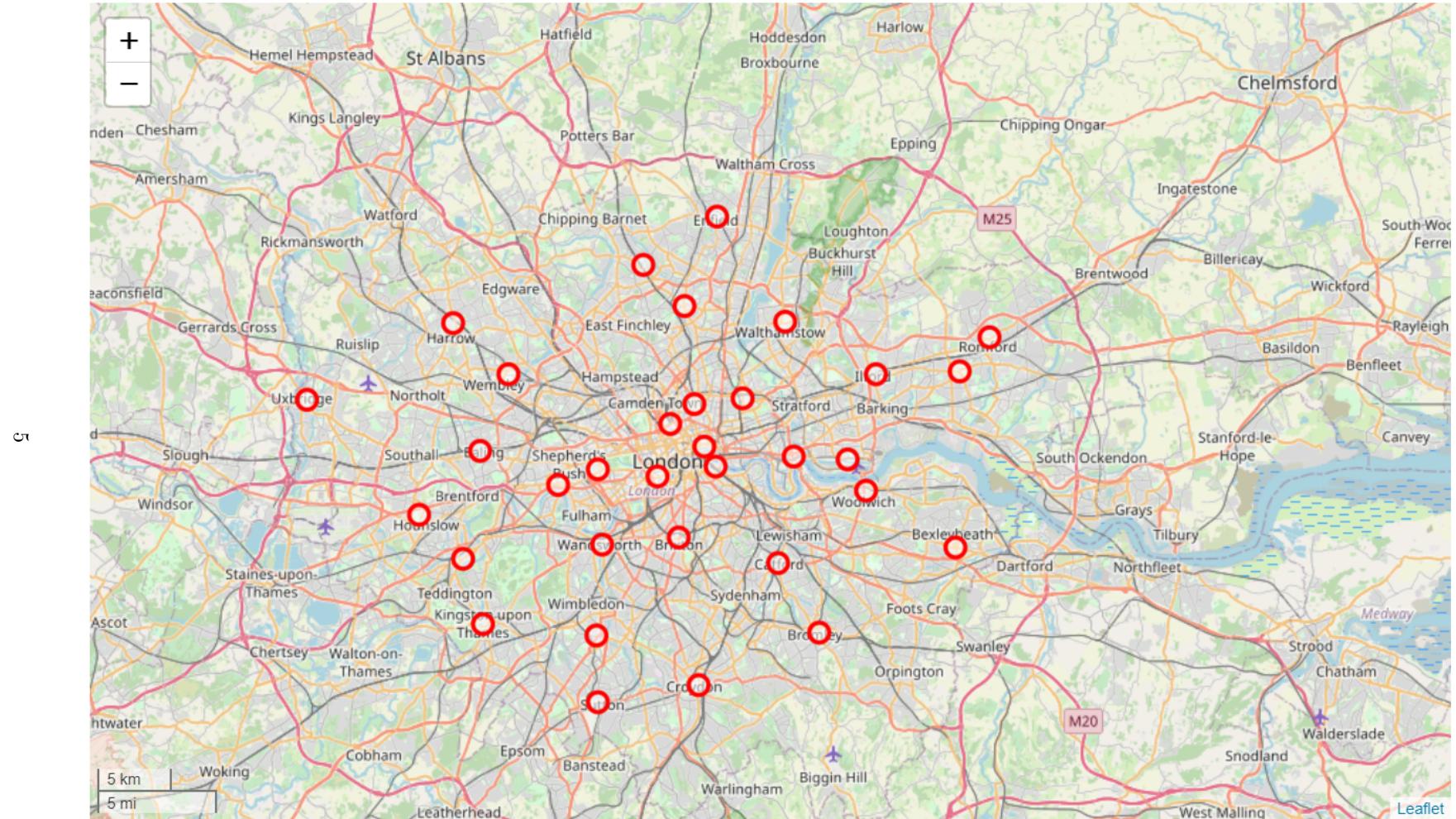


Figure 2: Map of all London Boroughs

4 Results

Firstly, it must be noticed that frequencies will now be expressed as percentage values as it might result more clear for the reader. Furthermore, the results for Inner and Outer London will initially be shown in two different sections, with the final section dedicated to the overall view of all the data.

4.1 Inner London

Table 3 shows the *IRF* values within the Inner London boundaries. It is immediately visible that the two areas with the highest competition in terms of Italian restaurants are Lewisham and Hammersmith and Fulham. Both these Boroughs have an *IRF* of 8.00% - by far the highest of the lot. On the other hand, the lowest Italian Restaurant Frequency is found in Newham with *IRF* = 1.82%. Newham is closely followed by Haringey, with an *IRF* of 2.00%.

Table 4 displays the Restaurant Frequency in each Inner London Borough. Once again, Hammersmith and Fulham ($RF = 39.00\%$) is amongst the areas with the highest frequencies together with Kensington and Chelsea ($RF = 37.00\%$) and with Lambeth (the overall highest $RF = 40.23\%$). Just like for the *IRF*, Newham shows the lowest value of *RF* with 16.36%. This value is the lowest by a safe margin, as Wandsworth is in second position with $RF = 20.45\%$.

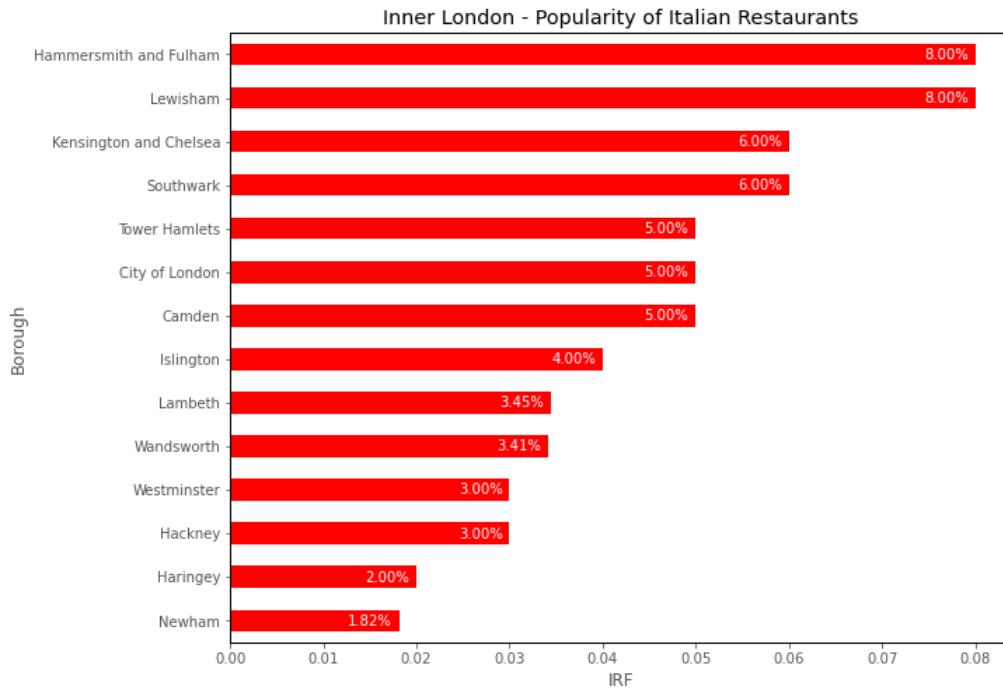


Figure 3: *IRF* for Inner London Boroughs

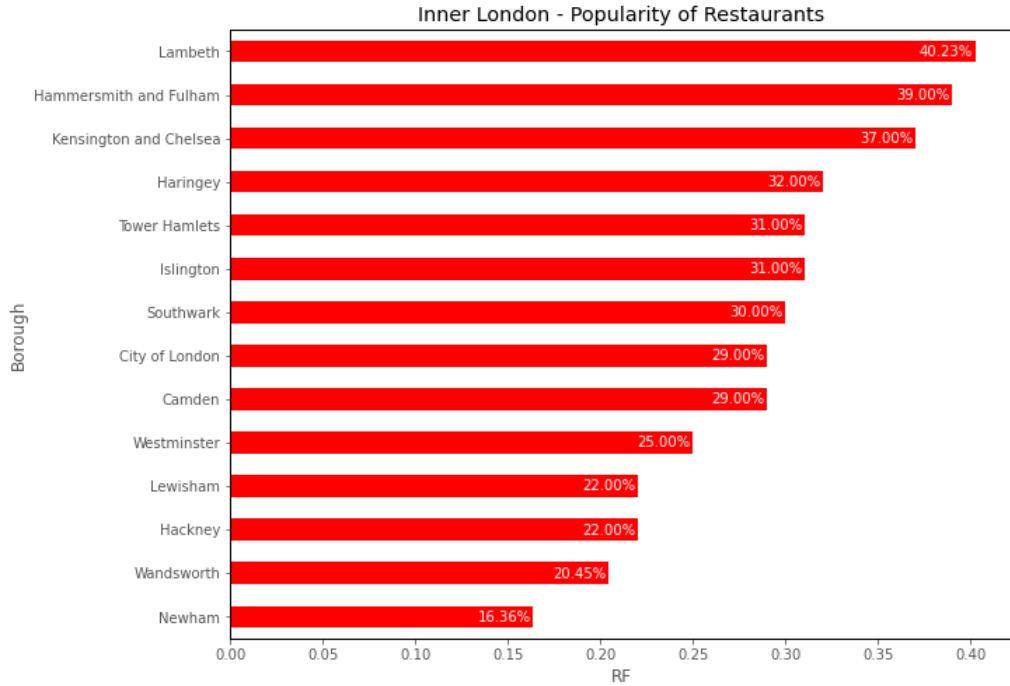


Figure 4: *RF* for Inner London Boroughs

4.2 Outer London

The Outer London *IRF* data are shown in table 5. Hillingdon (*IRF* = 1.79%) has the lowest frequency of Italian restaurants, followed closely by Hounslow (*IRF* = 2.17%), Kingston upon Thames (*IRF* = 2.22%) and Croydon (*IRF* = 2.50%). All the others do not show comparable values, as they all equal or exceed 4.00%.

The Restaurant Frequency among the Outer London Borough is shown in table 6. The lowest value is displayed by Barking and Dagenham with *RF* = 14.93%. The second place is taken by Greenwich with *RF* = 15.15%, followed by Havering *RF* = 16.33%.

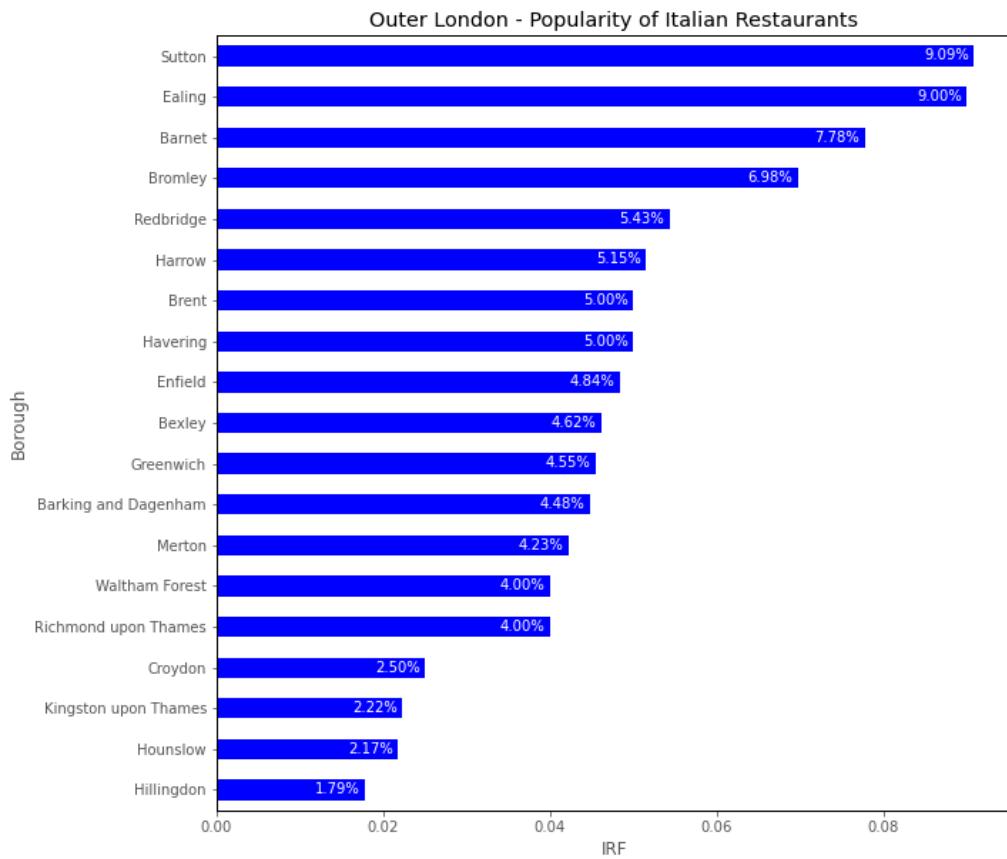


Figure 5: *IRF* for Outer London Boroughs

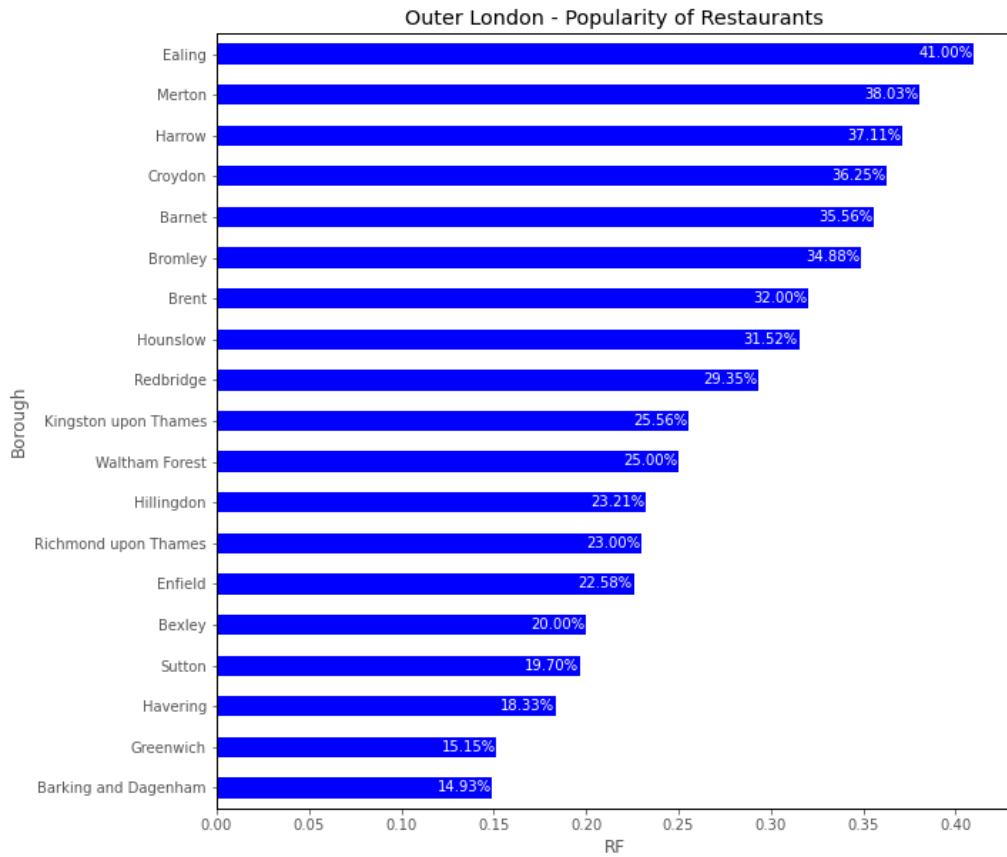


Figure 6: *RF* for Outer London Boroughs

4.3 Overall Analysis

From the Inner London plots in section 4.1, it is extremely easy to identify Newham as the best possible Inner Borough, since it has the both the lowest *RF* and *IRF* in its category. However, it might not be as simple for the Outer London Boroughs. In terms of *IRF* the lowest 4 Boroughs had a good margin on the others. However, they do not have an extremely competitive *RF*, as none of them is within the lowest 7 positions in table 6.

This is where a data visualization method that can include both frequencies can help us analyze the results. For example, a scatter plot with the two frequencies along the two axis is displayed in figure 7. This graph also allows to include all the 33 Boroughs in the same plot, without the need to divide them into Outer and Inner London.

The scatter points that are closest to the origin are related to the Boroughs that have the best combination *RF* and *IRF*. Newham has the second best *IRF* (1.82%) which is comparable to the best value overall (Hillingdon with 1.79%). At the same time it has the third best *RF* (16.36%), with Greenwich and Barking and Dagenham with respectively *RF* = 15.15% and *RF* = 14.93%).

5 Conclusion

The reported data clearly show that Newham is the best possible Borough to open a new Italian restaurant in London, being the most competitive combination of *RF* and *IRF*. Even though Newham is located on the external part of the Inner London, it is an area of great economical interest being home to the London City Airport. Other areas that are worth being taken into consideration in further business analysis are Greenwich and Barking and Dagenham, where the presence of restaurants among the most popular locations is the lowest overall (lowest *RF*). Interestingly, these two Boroughs are adjacent to Newham, making the whole East London a very attractive area for new restaurant entrepreneurs. It is worth mentioning Hillingdon (lowest *IRF*) where the availability of popular Italian restaurant is the lowest. All these Boroughs are displayed in the map 8 (interactive version available in the code provided).

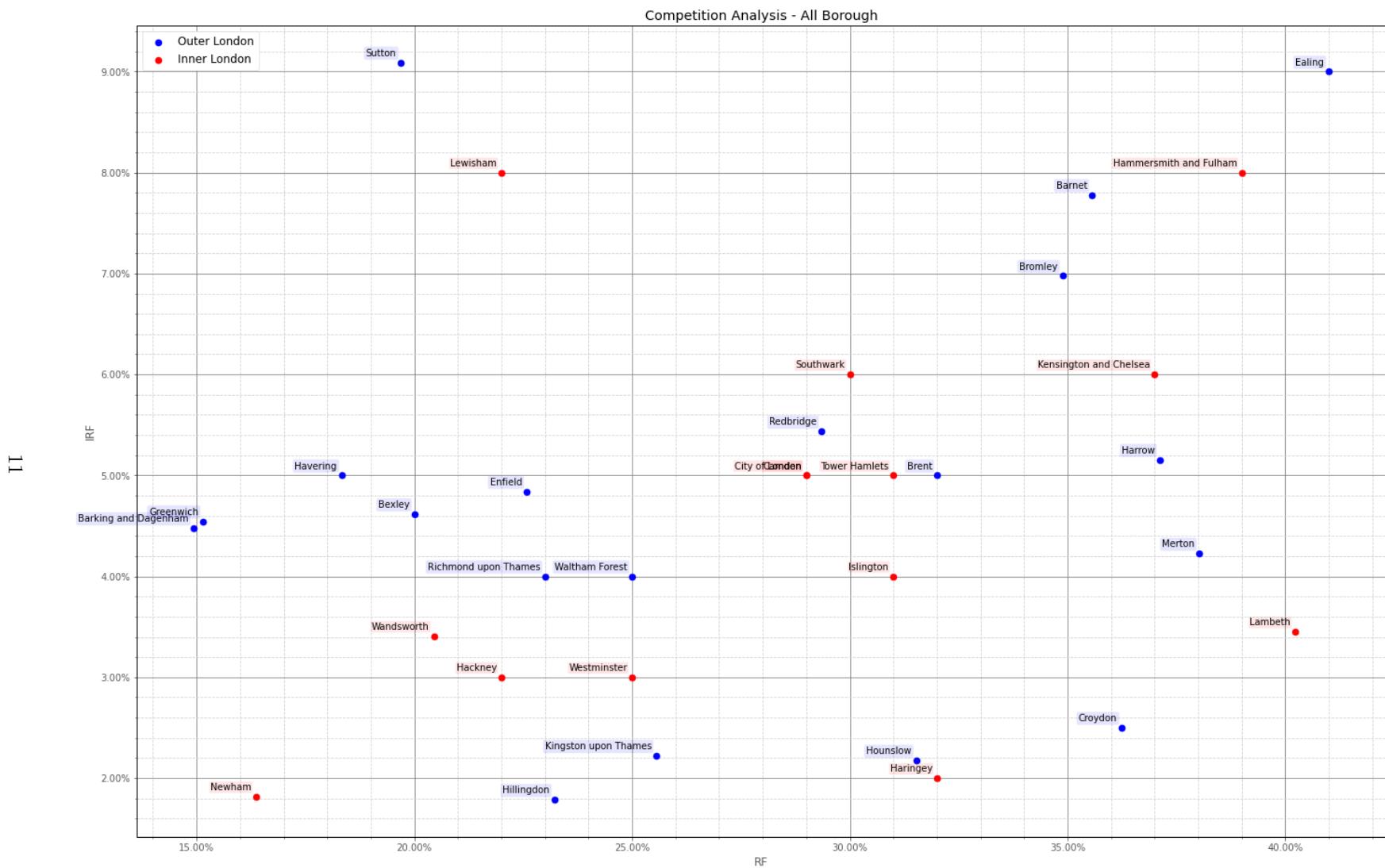


Figure 7: RF and IRF scatter plot for all Boroughs

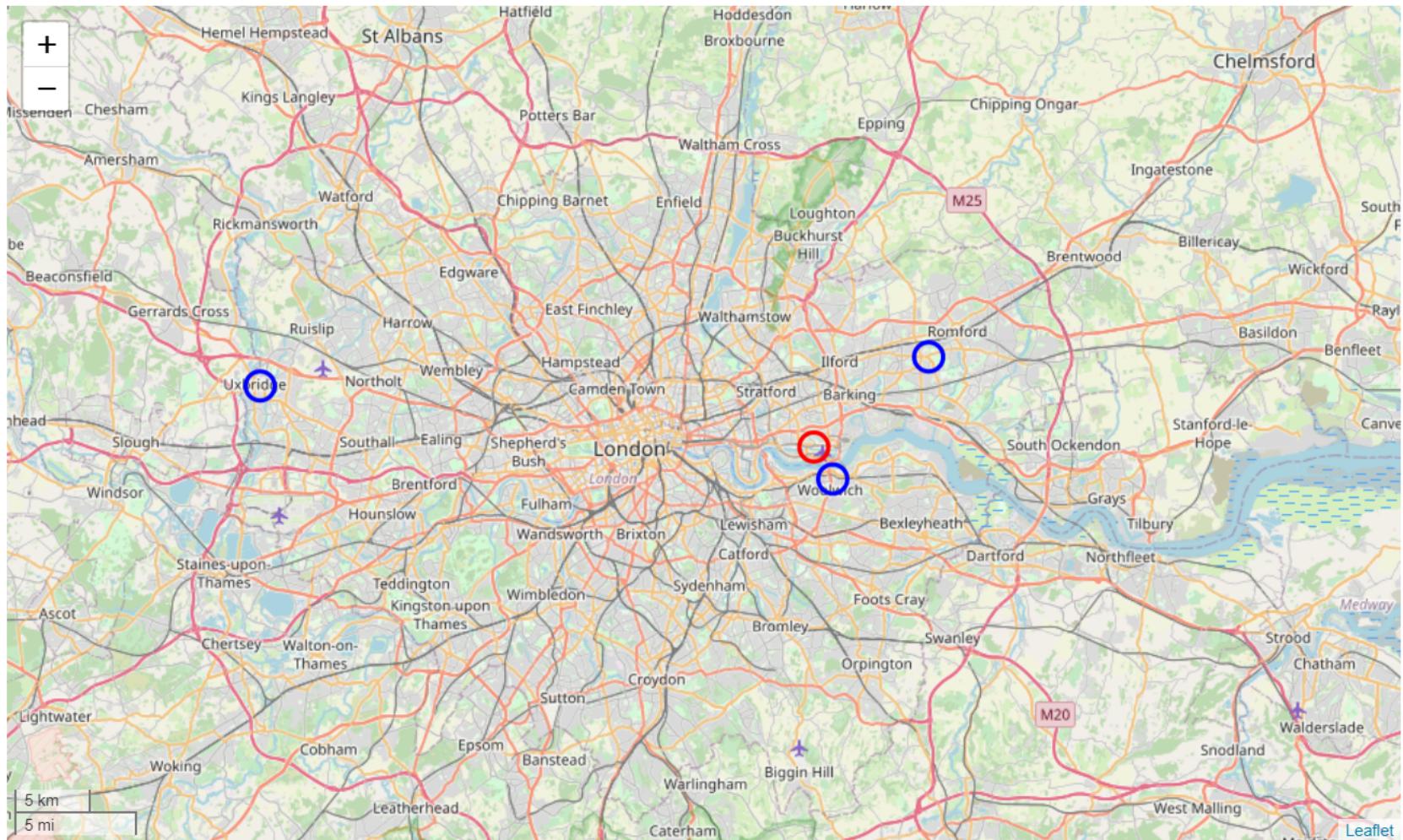


Figure 8: Map of the best London Boroughs to open a new Italian restaurant. Inner Boroughs are displayed in red, while outer are blue

References

- [1] Wikipedia. List of london boroughs. https://en.wikipedia.org/wiki/List_of_London_boroughs. Accessed on 2021-01-02.
- [2] Wikipedia. Inner london. https://en.wikipedia.org/wiki/Inner_London. Accessed on 2021-01-02.
- [3] Wikipedia. Demography of london. https://en.wikipedia.org/wiki/Demography_of_London. Accessed on 2021-01-02.
- [4] Coursera. Ibm data science professional certificate. Applied Data Science Capstone.