

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 9, 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 reflects 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 of this report would be 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 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.

Finally, it is important to discuss the two main parameters that will be 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 with 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:

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

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

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.