**Finding Optimum Locations to Open a Restaurant in London**

Samer Bachir

10th November 2020

# Introduction

## Background and History of London Restaurant Business

London's restaurant industry started at around the end of the 18th century, where Londoners would eat out at inns and taverns from where the restaurant industry initially spawned from. Some of these establishments still exist today, such as the Wiltons which opened in 1742 selling mainly seafood in Haymarket. At around the 1860s the cuisine styles being served began to vary as the UK became more industrialised. In 1867 a former chef to Napoleon III opened a French restaurant in London. Ever since, the UK's trade with the commonwealth and the rest of the world resulted in Londoners being treated to many different cuisines and experiences. Trends in migration to the UK only increased the quantity and variety of restaurants in the capital until today it is awash with restaurants from every part of the globe and Londoners frequently make it part of their routine to sample the interesting and exotic foods on offer.

## London Restaurant Trends

Neglecting the short-term impacts from COVID-19, the restaurant business in London has been growing at a good rate over the past 30 years. The number of restaurants now number almost 20,000, however there have been and continue to be significant changes to the makeup of restaurants in the capital. In more recent times these seem to be mainly driven by demographic and technological changes.

Millennials are a big driver in change in the restaurant business. They tend to value individuality, uniqueness and adventure. They are also less loyal and more likely to follow the crowd to the latest opening for the social media kudos. Millennials are generally considered to have been born between 1981 and 1996 and so in 2020, are aged between 24 and 39. For this generation dining out is an experience to share with friends and meet new people. They choose restaurants for the atmosphere and enjoy engaging with the creativity and character of the restaurant. As more and more of these millennials become parents, the trend for kid friendly dining experiences is set to increase. This generation has grown up eating out more often than their parents and will take their children out with them and expect more than the standard and typical children’s menu of yesteryear. Health and sustainability are becoming increasingly important factors in the industry as society shifts its focus onto these important global scale problems. It is expected that regenerative agriculture, organic sourcing of ingredients, more vegan and vegetarian options and awareness of reducing food waste will become important to restaurant goers and they are prepared to pay more for these options.

The relevance of the above to this assignment is just to highlight that many of the factors that made a restaurant successful in the past are not likely to be the same as those that make a restaurant successful in the future. Using machine learning techniques such as regression analysis that looks at past data to discover relationships from the data have to be done so carefully otherwise all you will obtain is a successful restaurant of the past, and perhaps not a successful one in the future.

## Restaurant Business Problem

The business problem I have decided to solve is to determine the best location to open a vegan or vegetarian restaurant in London. It is expected that this type of restaurant will mainly appeal to the millennial generation, although there is a growing trend within the older generations to be more sustainable and healthy as well. I will use a variety of data sources to try and select suitable locations that well suit the millennial generation.

The restaurant business is incredibly competitive in central London, just serving good food is rarely enough these days to avoid failure. Nowsourcing estimates that almost 60% of independent restaurants fail within 3 years. Some of the main factors for failures are as follows according to The Career Cafe:

1. High concentration of restaurants in a given locality and saturated market.
2. Owners skilled at cooking but not at business.
3. Owners inability to raise capital.
4. Owners inability to conduct proper marketing.
5. Locating a restaurant in the wrong demographic or where there isn't enough footfall.
6. Owners having to work too many hours and having not the quality of life they originally planned for.
7. Not providing a full customer experience that appeals to all the senses.
8. Lack of differentiation in brand concept with other nearby restaurants.
9. Paying too high a rent for the revenues that could be reasonably achieved.

Given that the business problem to solve is strictly around selecting an appropriate location, only the following from the above factors are relevant: 1) A high concentration of restaurants in a given locality. 2) Locating a restaurant which is in the wrong demographic. 3) Locating a restaurant where there isn't enough foot fall. 4) Lacking differentiation in the brand concept with other restaurants in vicinity. 5) Paying too high a rent.

As well as avoiding common reasons for failure a restaurant should make the most of the following factors that can affect how much revenue a restaurant can make according to RestoHub:

1. Having excellent transport links.
2. Being in a location where your target market make a good income.
3. Being in a location where there are many complementary businesses for restaurants, such as; Shops, bars and theatres.

## Restaurant Stakeholder Perspective

The target audience for this work are for individuals looking to open a vegan or vegetarian restaurant or companies that provide financing to this segment of the restaurant industry. This work will enable the above groups to make important data backed decisions on where a good location may be. Given the failure rates within the industry and that location is a key consideration in having a successful restaurant business, the above information will give the individuals a good sense if a prospective location is going to help or hinder the prospects of the future restaurant.

Given that London is such a large city and with such variety, conducting any type of manual exercise would be too much effort and take up too much time. Furthermore, using individuals who may be well experienced in the restaurant business may not lead to a positive outcome given the amount of change that is currently going on. This is more so given the target demographic group and macro-economic trends that are taking place.

# Data

## Filtering Down the Data Sets

As discussed above it is suggested there are eight criteria that may be significant in selecting the most appropriate location for a restaurant. Given the scope, intent, and the lack of any financial budget for this study, it is not considered appropriate to conduct this study on so many different variables. Therefore, the above criteria will be reduced to only those considered to be the most impactful and those available in the public domain. Following this reduction exercise the following location related factors have been selected:

1. Excellent transport links.
2. Not having a high concentration of competing restaurants nearby.
3. Having many complementary businesses nearby.
4. Having high local population densities nearby.

The other criteria were not selected for the following reasons:

1. Locating a restaurant that is in the wrong demographic was not selected in the final list since it was not possible to find accurate information on this from a free data set. Also, selecting the right demographic area is actually very complicated, since millennials tend to live in one area and work in another and go recreate in perhaps an entirely different one.
2. Locating a restaurant with enough footfall was simplified by estimating the population density in the vicinity of proposed restaurant locations. Foursquare analytics do provide detailed footfall information in London but this is not a free service. An example of the output that can be obtained from Foursquare analytics can be seen in Figure 6.
3. Lacking brand differentiation was simplified to just evaluating if there are restaurants in the vicinity that were of the same type. Clearly, there is more to brand differentiation to just this, but it would not be practical to try and put together the data that this would entail given how time consuming this would be.
4. Rental information and local income information was not selected since all the free data sets that could be found on this were at the borough level and not at sufficient granularity to be useful. In other words, it was not really possible to calculate what the rent would be at a given location, nor what the average income would be in the vicinity of that location of my target market. In some ways this information may work to cancel themselves out since in London, typically where you have a high rent you would also have a local population with a good income.

Given the above inaccuracies, it would be recommended that the locations which are selected by this evaluation are studied further to determine the impacts of the above factors and ensure a suitable location is selected against all the important criteria.

## Data Sources

For the final four criteria that have been selected, the required data for them will be sourced as follows:

1. Excellent Transport Links: Excellent transport links shall mean how close the restaurant is to London underground and train stations. For London underground and train information a dataset from Transport for London has been sourced which has the latitude and longitude for each underground and train station.
2. Concentration of Competing Restaurants in Vicinity: For this information, data from the FourSquare API shall be called. Latitudes and Longitudes will be provided by the API and these will be inputted into a Dataframe. The closer a prospective restaurant is to one or more restaurants of the same type the less the algorithm will select this location.
3. Concentration of Complementary Businesses in Vicinity: For this information, data from the FourSquare API shall be called from. Latitudes and Longitudes will be provided by the API and these will be inputted into a Dataframe. The closer a prospective restaurant is to one or more complementary businesses the more the algorithm will select this location. In order to not allow this assessment to be too exhaustive without any significant increase in accuracy, only bars and theatre locations were included to represent the general factor of complementary businesses. This is probably not too an inaccurate simplification since where there is a significant concentration of bars and theatres it would also be expected that the concentration of other complementary businesses would also be high.
4. Local Population Density: A dataset is provided on the UK's Office for National Statistics (ONS) that provides population information for each borough in London. This format will have to be translated into one compatible with a map. Fortunately, this dataset also provides more detailed population density data down to mini-areas called ‘Lower Layer Super Output Area’ (LLSOA). These mini-areas in London the ONS provides also include each area's latitude and longitude which will be simple to convert for use into this assessment.

Clearly, to select good locations more accurately for the prospective restaurant it would be important to understand the relative importance of the above factors to determine which factors have more effect on success than others. Furthermore, it would be important to understand the relative strength for each criteria. For example, how the proximity to good transport links and complementary businesses alters with increasing distance. In other words, if a train station were 10 minutes walk away at one location and 20 minutes away from another location, how much better would this be for the closer restaurant? Ideally, this type of analysis lends itself well to machine learning techniques such as a regression analysis to help select the weightings on these criteria. These methods however need a value that would accurately gauge how successful a restaurant was. That way the model or algorithm could be accurately trained and then tested. However, finding such a dependant variable is not straight forward and then finding an accurate dataset for said variable would be even more challenging. If this problem were to actually be requested by a client, it is recommended that detailed footfall information is purchased from foursquare over a recent historical period and then use this as a dependant variable.

## Data Wrangling

An issue that frequently presented itself when obtaining data from the from the Foursquare API was an inaccuracy in the category results that were passed back. Although the reason for this error is unclear, it was remarkably consistent and random in terms of what errors were occurring. A check was therefore always conducted by using the ‘str.contains’ method to ensure only the values that were desired were obtained.

One of the early decisions that was to be made was over the use of which coordinate system to use. There are several different systems that are used, however, the two most pertinent to this work are the Geographic Coordinate System (GCS) that uses longitudes and latitudes to give positions of objects on the Earth’s surface and the Cartesian System that uses a horizontal and vertical axis from a particular reference point. Most data that can be found in public domains or called from APIs are presented in the form of GCS coordinates, however, to be able to use grids over an area of London, the Cartesian System is more effective. This resulted in frequently changing between one system and the other and the mathematics in order to do this is far from simple. However, by installing a couple of modules that are designed to handle such conversions, this process was made to be much faster and simpler.

Frequently when using APIs, a limit is introduced to ensure the API does not undergo excessive workload and use up too much computing resources. In the case of the Foursquare API a limit exists where the number of output results that could be obtained in one call was 100 items. Given the number of restaurants and complementary businesses there are in central London, 100 items was far too small. In order to circumvent this problem the London was area was divided into 2.5km radius circles and a call to the API was made for each of these areas. The outputs were then accumulated into a separate Dataframe for further processing.

# Methodology

## Exploratory Data Analysis

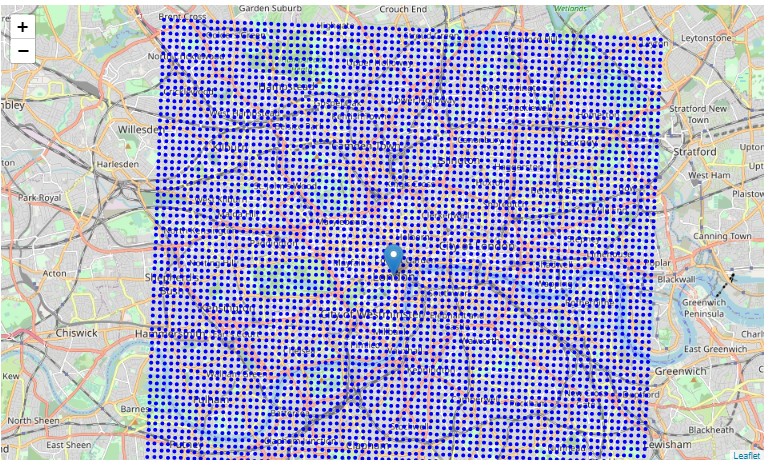
Exploratory data analysis was conducted on the population density factor. The number of Lower Layer Super Output Areas was significant, almost 5000 individual areas. However, many of these areas were some distance away from the central London pixel grid and therefore would not be relevant to the assessment given their limited affect. The distances and values were assessed against the impact they would have to the score and values were dropped out if they were not significant. This assessment resulted in the number of LLSOAs to reduce by over half and therefore significantly reduce the amount of processing that was required to be undertaken. Figure 1 below shows the population density versus the LLSOAs, the darker shade of green the higher the population density.



Figure 1 – ONS Data on Population Density in each London Lower Layer Order Super Area

As mentioned briefly in a previous section, the relative relationship between each factor and how they affect the final location is a significant aspect to this assessment. How much more (or less) important is being near a train station is it than being near a theatre? Without being able to use machine learning techniques there was no accurate way to answer this. For the purposes of this assessment, the assumption made that each factor was around as important as each other. In order to manipulate the numbers to ensure they have equal weighting, the ‘describe()’ method was used to find the mean and distribution of each dataset and the weightings adjusted accordingly.

One of the decisions that needed some trial and error was what the appropriate size for each pixel should be. Several different sizes were trialled so as to determine what the impact would be to the granularity of the visual output of the results. It was found that using large pixels of around 0.5x0.5km and some dummy results not enough granularity resulted in the output in order to give a reasonable output. When using small pixels of around 100mx100m the time to undertake processes (particularly the coordinate conversions) was prohibitively long. Finally, a 200mx200m pixel area was selected as it was considered to strike the best balance between having enough granularity to show good restaurant locations but at the same time not having a long run time to perform the necessary calculations. Figure 2 below shows what the 200mx200m grid looked like when it is superimposed on top of the London map. There are 75x75 points which represents an area of 15kmx15km.



*Figure 2 – ONS Data on Population Density in each London Lower Layer Order Super Area*

## Statistical Testing

Figure 3 below shows box plots that allow for quick evaluation of the statistical distribution and main statistical properties of the weightings used in the assessment. The first five box plots show the statistical summary of the factors that were modified to ensure that all five factors have a relatively equal weighting. The last boxplot is simply the sum of the five values. Please note that the box plot titled “Veg. Score” is a negative number where the mean is aimed to be around -10. Although the weightings were successfully modified to produce reasonably equal weightings, there were a large number of very large positive and negative values that were obtained with this assessment. This could be due to some values being extremely close to the location of the pixel centre point and skewing the result.

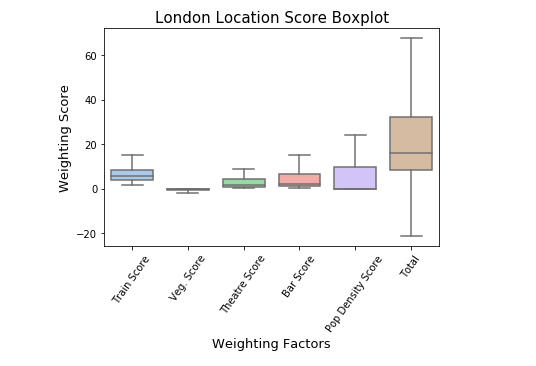


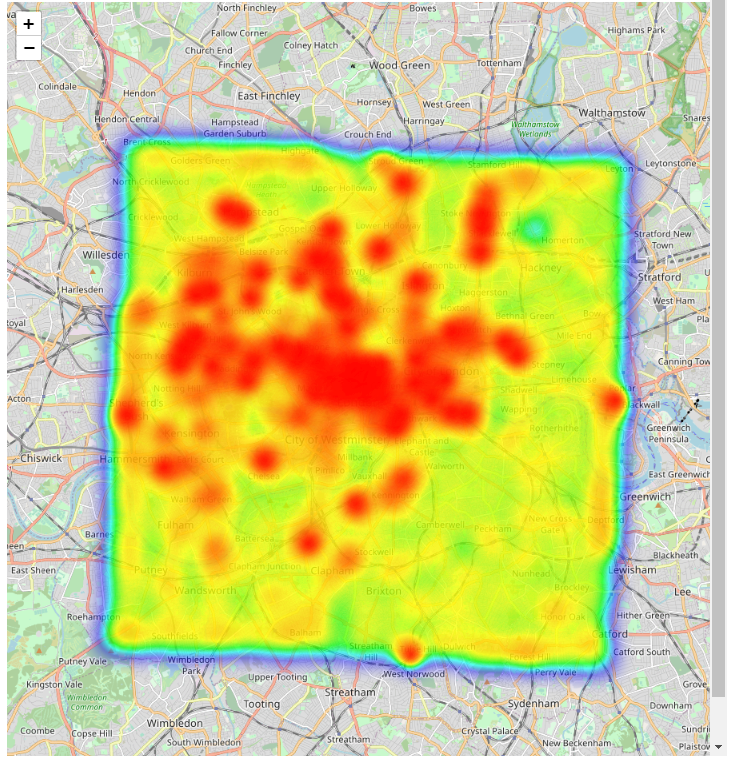
Figure 3 – Box Plots Showing Distribution and Means for All Weighting Factors

## Machine Learning

As mentioned in a previous section, being able to use machine learning techniques would have very useful in this type of assessment. Having so many weightings and factors to modify would have meant that further understanding into the relative importance of factors could have been obtained. Furthermore, it would have been possible to work out what the expected accuracy of the final algorithm would have been and provide any prospective client more confidence in the outcome. It is recommended in future work to undertake machine learning techniques such as regression analysis and possibly neural network based methods to obtain a more accurate and testable algorithm.

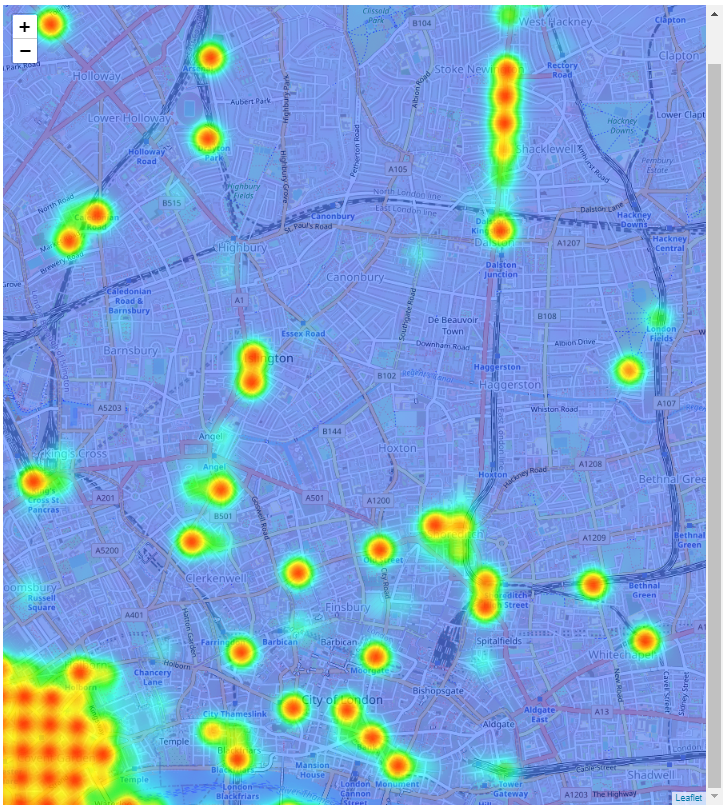
# Results

After a total score was obtained that reflected how good each potential restaurant location was, this data could be transferred from a Dataframe into a plot. It was considered that a Heat Map would be the best method to visualise this information and the Folium package includes such a map method. The parameters it takes are relatively straightforward – longitudes, latitudes and weightings. There are some additional parameters that affect the look of the map which require some trial and effort in order to optimise. Figure 4 below shows the overall Heat Map for the central London area, the warmer colours represent the more optimal locations.



*Figure 4 – Heat Map Showing Best Vegan Restaurant Locations in London*

The Heat Map is interactive and is easy to zoom into and enables the user to obtain a more detailed and granular view into what is happening in areas that are of particular interest. Figure 5 below shows a northern slice of the above map in a lot more detail. It can be seen that favourable restaurant locations are located close to stations and areas of town near many attractions, particularly near current high streets. It is also worth noting that the population density seems to have a more broader and less spiked effect on the results.



*Figure 5 – Heat Map Showing Best Vegan Restaurant Locations in North London*

# Discussion

## Observations

The results that have been summarised using the Heat Map in Figure 5 show which locations would be good to open a vegan restaurant and which would not be. Therefore, as long as the assumptions made in this report are accurate, it can be seen that the central parts of London are particularly attractive locations. This is not surprising given the number of train stations and complementary businesses that are situated in this area. Unfortunately, this assessment does not factor in the rental expense, as this is significant in central London and will affect the economics of any prospective restaurant location. Further to this, it can be seen that the West part of London seems to have more hotspots than the East and South parts of London, this is probably related to the number of complementary businesses that concentrate in this part of London.

Interestingly there are a few hotspots that are not so near to central London and likely to have cheaper rental values. Depending on type and budget of restaurant that will be opened, these areas should be assessed in greater detail. Figure 6 below shows a download from the FourSquare Checkin app which shows what people in London are doing at various times of the day. This and similar data that FourSquare store would be very useful to this assessment.



Figure 6 – Foursquare Analytics Visual Data Output over Time of Day in London

## Recommendations

1. Given the above inaccuracies, it is recommended that the locations selected from this evaluation are studied further to determine the impacts of the above factors and ensure a suitable location is selected against all the important criteria.
2. It is recommended that detailed footfall information is purchased from Foursquare over a recent historical period and this dataset used as the dependant variable on which to base this assessment.
3. It is recommended in future work to utilise machine learning techniques such as regression analysis and possibly neural network based methods so as to obtain a more accurate and testable algorithm.
4. The rental expense of individual areas should be factored into a future assessment since this assessment seems to be fairly biased towards the most central parts of London where rental prices are expected to be very high.
5. Further assessment of areas away from central that have been assessed as optimal should be looked at in further detail to determine if there are potential good opportunities at relatively low cost for prospective clients.

# Conclusion

Notwithstanding to coarseness of the above assessment, it would still be very useful to be able to test how accurate the model is. There are several ways in which this could be done, for example, the profitability of restaurants in different areas, how long restaurants survive in given areas, or what the footfall is at different restaurants. Unfortunately, there does not appear to be any straightforward way to obtain any of these datasets. So, although, the logic and criteria that have been selected appear sound, there is no real way to test it, which would also make this assessment difficult to sell to prospective clients.

The above assessment would also greatly benefit in the use of machine learning techniques such as regression analysis and even more sophisticated approaches such as neural networks to more accurately weight and calibrate the factors that have been identified. This could even remove factors that are not effective in predicting the best locations and perhaps help discover other factors that may be more significant. However, without a dataset that can accurately ascertain how successful a given location is, none of these methods would work.

I believe the natural extension to this work would be obtaining more accurate data and use machine learning techniques as mentioned above. I also believe that more work should be done to better understand the habits and preferences of vegan restaurant goers. This will lead to perhaps more factors being identified that can find locations where perhaps typical restaurants would not be successful, but vegetarian/vegan restaurants would be.