Finding Untapped Regional Cuisines in London's Catering Industry

1 Introduction

London is the epicentre of diversity in the United Kingdom, which is reflected in the city's equally diverse catering industry; bringing a myriad of different global cuisines to the economic forefront, many of which being family-run businesses. This project looked to explore possible correlations between the borough demographics of London and the state of the the same borough's local catering industry. If the ethnic base of the borough shows to have an effect on the industry, then it may be possible to unveil areas which have untapped market potential.

This kind of business insight may be valuable to aspiring businesspersons and the corresponding landowners as they can see if the desired build location has the same makings as other successful restaurants elsewhere, giving much needed certainty in the starting process.

2 Data

The project centres around the combination and cross-analysis of three key datasets:

1. An abridged form of the London borough data from the Wikipedia page; outlining the coordinates, area, and population.

2. The foursquare API to corroborate with the location data, in order to provide data for the venues in the vicinity. This dataset will be at the core of analytics.

3. Demographic data for each London borough which can be cross-referenced with the venues, to try to draw conclusions on the part demography plays on the industry.

The reason as to why the location data was split into boroughs instead of other forms of division (postal code, district etc) is that boroughs have slight autonomy over laws which may affect businesses. Thus from a business perspective there is worth in deciding on a specific borough over a district which may be cross-borough.

Additionally the data to derive the demography of each London borough was taken from the Office for National Statistics (ONS), and instead of taking census demography data, the data used is based on the birthplace of mothers each year to give a more rigorous display of culture which may be hidden by simple ethnicity data.

3 Methodology

The general overview of the methodology was to build an intuition for the demographic distribution across the whole of London by first exploring the ONS data, then by using the borough data, invoke the foursquare API to further investigate how the catering industry changes from borough to borough.

This intuition was built by first looking at the demographic breakdown of each of the boroughs and also looking at the most multi-cultural boroughs, as well as trying to find the partition between inner and outer London, identifying any further differences between the macro groups.

Afterwards, the primary data set for boroughs must be cleaned from the raw Wikipedia table. This will include removing the political domains and centering the coordinates. Once cleaned the data can be used to invoke the Foursquare API, and the area column in the borough data can be used such that the foursquare pull request can have a dynamic radius to cover the whole borough. The radius input in the pull request will come from the area of a borough assuming that area is strictly circular.

Finally, the amalgamated list of venues from all the boroughs can be processed into a machine learning algorithm. To do this, it is first manipulated into a one-hot encoded form and then analysed using k-Means clustering. In this case, k-Means is the optimal approach as there is no need for a dendrogram and given that the area of each borough is already outlined, DBSCAN is not needed.

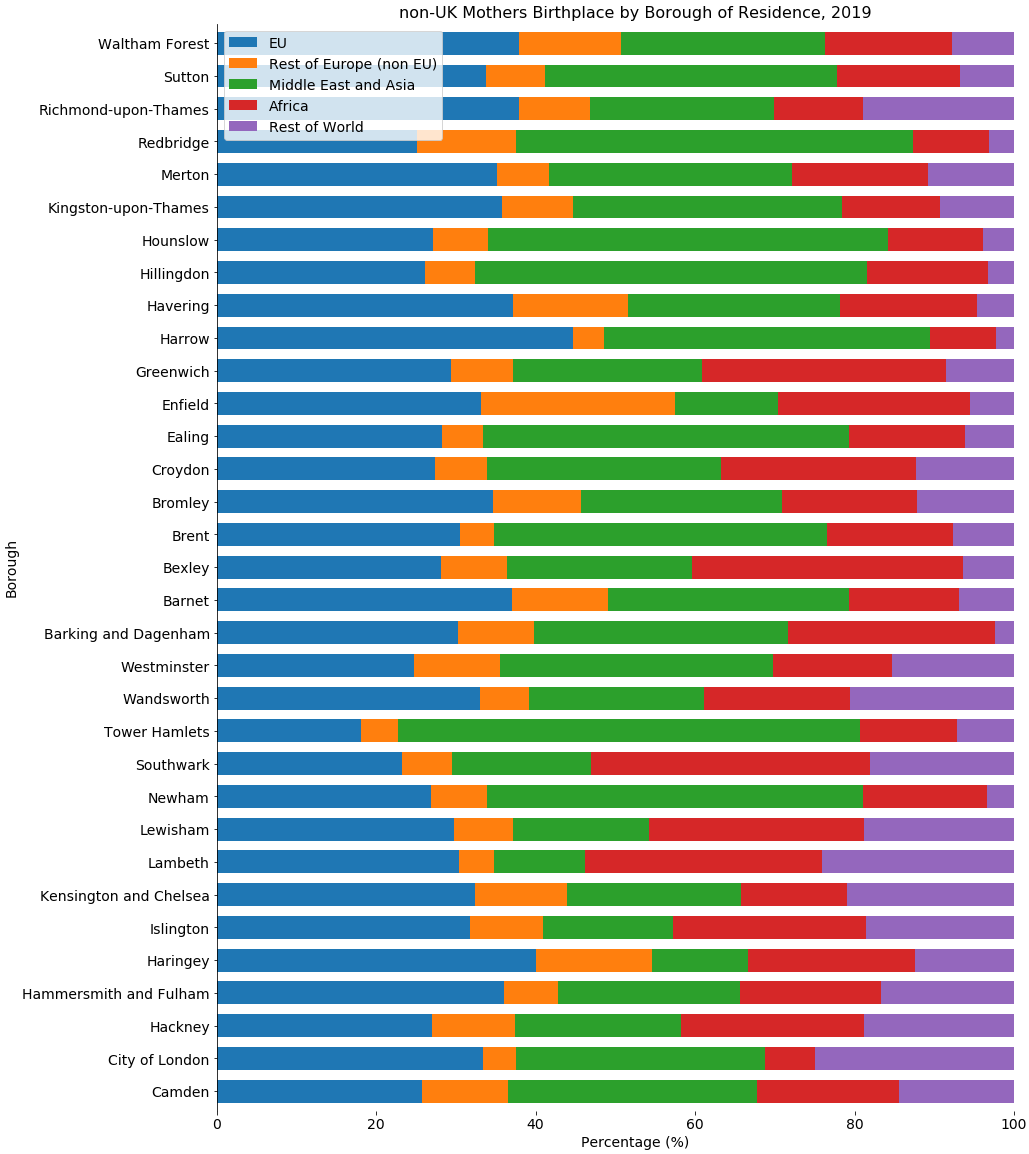

Figure 1: Stacked bar graph showing the breakdown of non-UK mothers birthplaces by borough of residence (including city of London) in the year of 2019.

4 Results

Th exploratory analysis of the demography data kicked off by simply looking at the ethnicity profiles of each of the boroughs to get a rough idea of the distribution. The result of this can be seen in the stacked bar chart in *Figure 1*.

The data described in *Figure 1* is the breakdown not including the UK-born mothers however it is possible that the numbers for EU-born mothers is inflated due to the large Irish population in England, which may not be useful for the investigation given how little Irish and British cuisine diverges. Despite that, we can see that Harrow has the largest population of EU-born mothers, Enfield has the largest population of Rest of Europe born mothers, Tower Hamlets has the largest population of Middle East or Asian born mothers, Southwark has the largest population of African-born mothers, whilst Lambeth has the largest population of mothers born from an uncategorised location in this study; this is likely made up predominantly of the Caribbean population.

We can also look how how multicultural each borough is generally, which can be simplified to the proportion of non-UK born mothers living in that residence. *Figure 2* shows a bar graph ranking the boroughs by this proportion.

Overall, we can see that London generally a very multicultural city; only three boroughs have less than 40% 2019 births from a UK-born mother, as well as 23 boroughs having over 50% of their 2019 births being from a non-UK born mother. From this point it was possible to try to draw a more specific relationship with the boroughs and their ethnic breakdown. My first thought was to see if there was any significant difference between inner and outer London, such that if one travels outward will the demography have a consistent change.

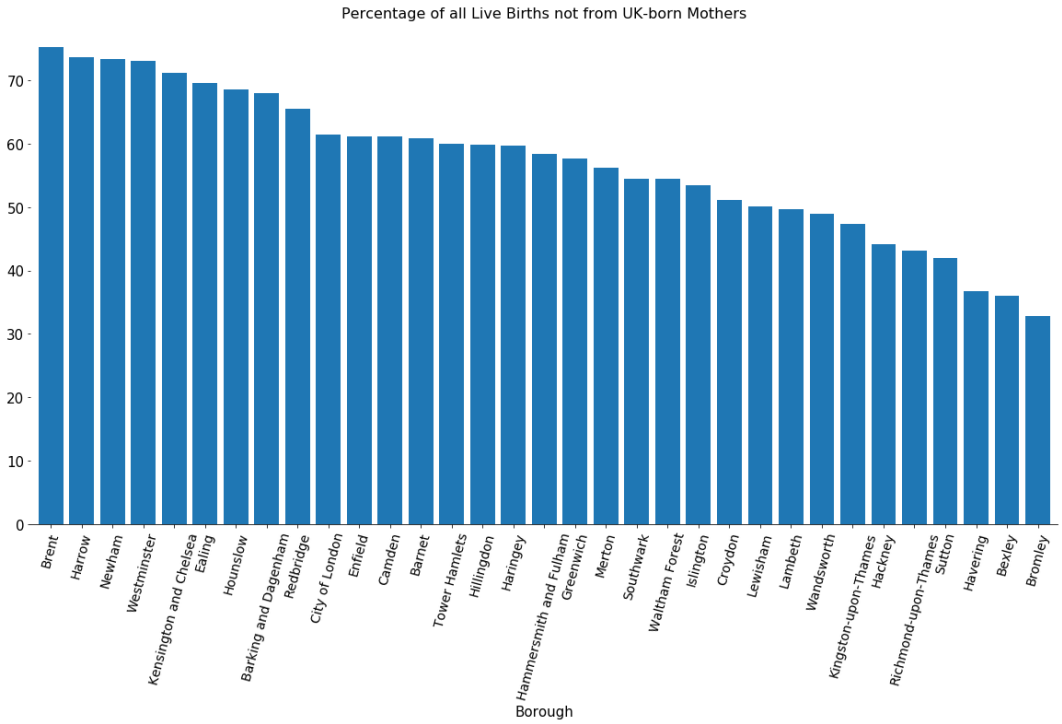

Figure 2: Bar graph showing the distribution of multiculturalism throughout London. Brent being the most and Bromley being the least.

According to the London Assembly, inner London includes the boroughs: Camden, Greenwich, Hackney, Hammersmith and Fulham, Islington, Kensington and Chelsea, Lambeth, Lewisham, Southwark, Tower Hamlets, Wandsworth, and Westminster. *Figure 3* shows pie charts of the demographic breakdown in the two subgroups. In the figure we can see that non-UK mothers in Inner London were mostly born in the EU, whilst in Outer London, this changes to the Middle East and Asia. However, given that the percentage delta is less than 5%, the inference can be seen as inconclusive for the scope of this investigation. Alternatively, we can note the large shift of the Rest of the World category, which shows the strength of the Caribbean community in Inner London.

Now, with the intuition built on London's demographics, it was possible to use the Foursquare API to pull the data on the locations of London's venues. For this investigation, we are looking at the catering industry in particular and so the venues were filtered by the second word 'Restaurant'. It is worth noting that some of food service venues had names such as 'Pizza Place' and thus were filtered from this dataset; given that those terms are mostly dominated by chain restaurants they often do not fully represent the cuisine that they are labelled as anyway. The subarray of just food service venues was collated and summed across the whole of London to provide the bar graph shown in *Figure 4*. this graph shows that out of all particular regional

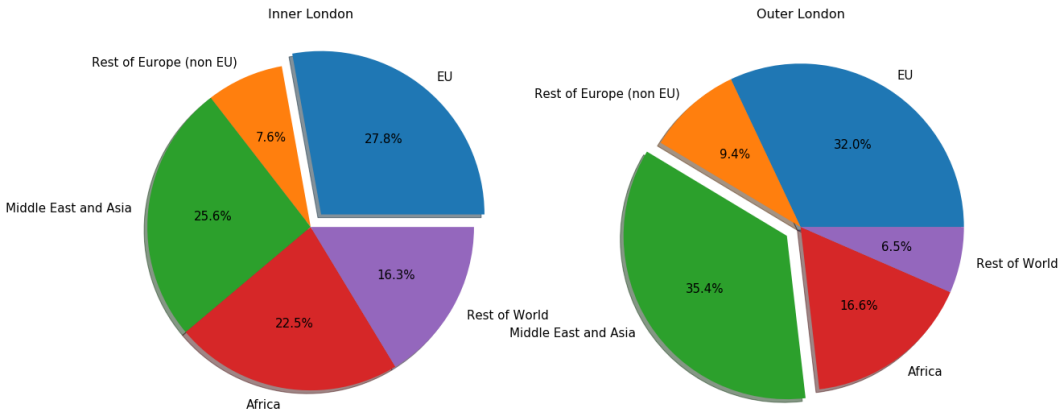

Figure 3: Two pie charts describing the summed demographic breakdown in inner (left) and outer (right) London. The exploded slice shows the the dominant demographic in each case.

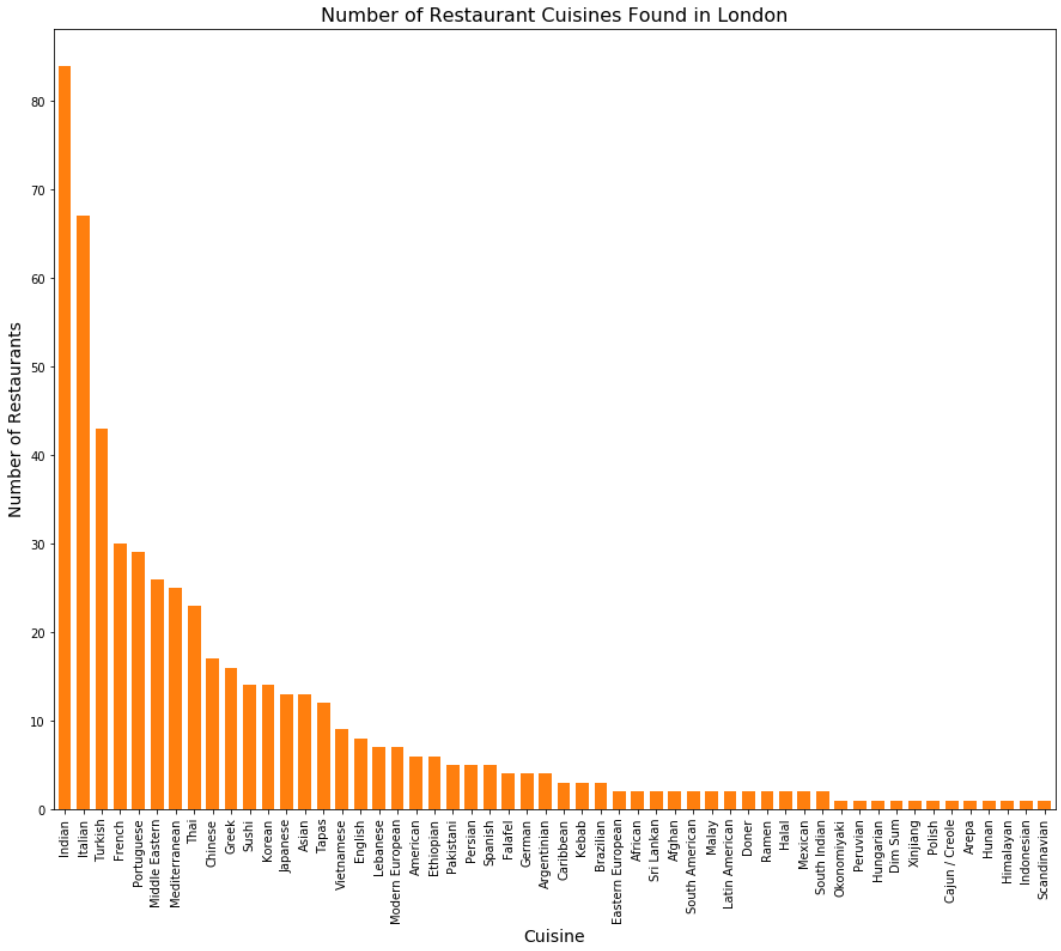

Figure 4: Bar graph displaying the number of restaurants which are listed as a unique cuisine. Some of these, however, not unique such as Sushi and Japanese restaurants.

cuisines, Indian and Italian cuisines dominate in London. Given the nature of the labelling from Foursquares API some of the cuisines are fragmented here, the best example being that 'Japanese' and 'Sushi' restaurants are seen as different however if they were merged, Japanese cuisine would be in the top five.

By converting the list of venues into a one-hot encoded format, we can process the data using k-Means clustering. For this investigation, six clusters were defined to give the best resolution, with 6 boroughs per cluster on average. The clusters centroids can be seen in *Figure 5*. Each cluster denotes a different set of boroughs which have a relationship with each other. For example the most common is shown by the purple labelled centroids, this cluster is defined by large quantity of pubs and supermarkets, and given that these are the two most common venues in London it makes sense that this is the largest cluster. However, the most noteworthy cluster for this investigation is the northernmost one in London, denoted with red labelling. This cluster includes Enfield, Barnet and Haringey, and is defined by the large quantity of Turkish Restaurants. This is the only cluster, when including all of the venues, that is defined by a regional cuisine due to a very high density of these restaurants. Looking at Enfield in *Figure 1*, we can see that it is the Borough with the largest quantity of mothers born from non-EU Europe, which of course includes Turkey, and when digging deeper Enfield is the hub of the Turkish community in London with a very significant proportion of the population made up of Turkish immigrants or citizens of Turkish descent.

It is clear from this that local ethnic pressure can have a very strong effect on the local industry, and even going so far to dominate the surrounding areas as seen by the cluster. This leaves the question of are there any other areas with this level of ethnic pressure but a vacuum for the corresponding local food industry. From our data we can detect a couple of candidates, the first and most crucial of which is Ealing in West London. This borough is the sixth most multicultural borough in London and has one the largest Middle Eastern and Asian populations, mainly comprised of Indians. Additionally whilst the percentage of Europeans is not the highest amongst the boroughs, the Polish population dominates in that category and makes it one of the boroughs with the highest Polish population. Despite these two observations, neither Indian nor Polish cuisines dominate here at all, even whilst the neighbouring borough, Hounslow, has 8% of all it's venues made up by Indian restaurants. It is therefore possible that Ealing is a largely untapped market for the global cuisine market.

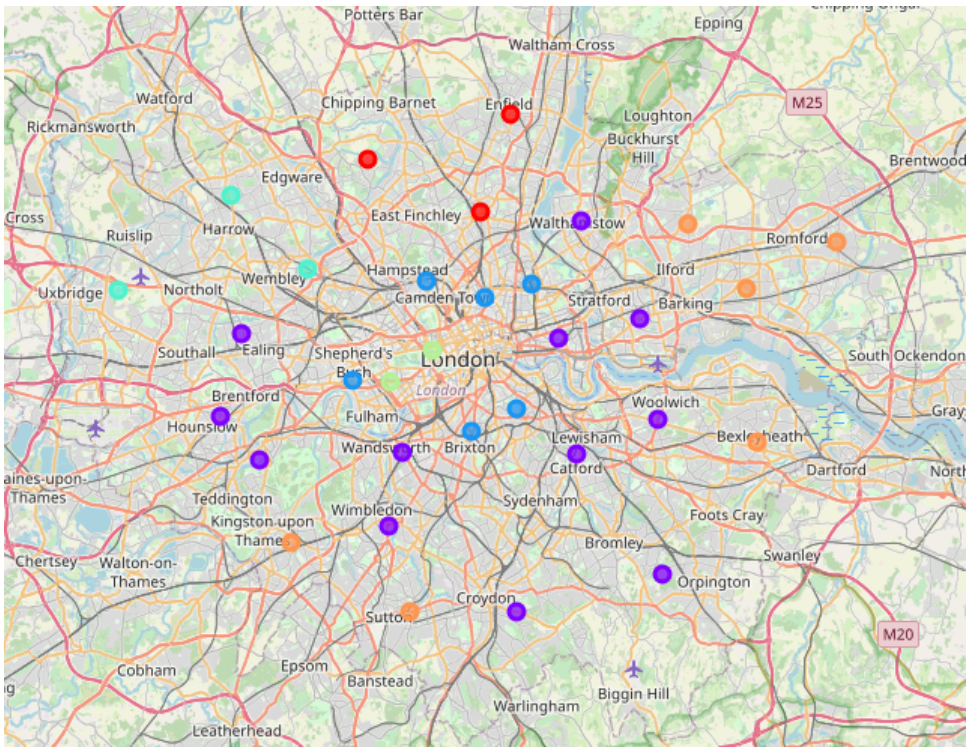

Figure 5: Caption

Another lesser case is that of Ethiopian restaurants, which may be undervalued in the borough of Greenwich. Greenwich has one of the largest proportions of African-born mothers, and a significant proportion of those are Ethiopian. This cuisine is generally unsaturated in London, with 84% of the Anglo-Ehtiopian population living in London, whilst Ethiopian cuisine only has 11% the presence of Indian cuisine. In Greenwich, there are zero Ethiopian restaurants despite a large part of the population being of Ethiopian-descent.

5 Discussion and Conclusion

Overall the investigation was able to draw some key conclusions in the attempt at finding regional cuisines in London's catering industry, in addition to wider conclusions. These were:

1. London's catering industry is mostly made up of pubs.
2. The demographic of local areas has a noticeable effect on the state of the local catering industry for the most part.
3. Using this knowledge, we were able to pinpoint areas (Ealing and Greenwich) which have cuisines (Polish and Ethiopian) which are wholly untapped.

Whilst the investigation was successful in terms of coming to a conclusion, it was not without flaw. One of two major pitfalls of the investigation was not being able to search in non-circular areas when invoking Foursquares API; the input only took a radius around a central coordinate, which means that for boroughs that are very irregularly shaped, a lot of venues were missed out on making the analysis marginally incomplete. The second was that the data provided by the ONS did have another level of breakdown of birthplace, only the region such as Africa or Middle East and Asia. This meant that I had to consult outside (non-governmental) sources about the breakdown of, say, the African population in Greenwich. As a result, some of the assumptions are made with concrete evidence.

Nevertheless, this was an insightful project and my first attempt at bringing together all of the course's points of interest together to make a cohesive end-product. i hope that these skills can be transferred further in the future.