Analysis of Indian Restaurants in and around Brick Lane, London

Problem Definition

Brick Lane is an historic street to the NW of central London, and dates back approximately 450 years. Once the scene of the crimes of Jack the Ripper murders, and home to the old Truman Brewery, it now has become a hub for the local artistic community featuring galleries, markets and festivals a throughout the year.

Being the centre of the local Bangladesh community, Brick Lane has gained the reputation as the centre of authentic Indian cuisine in London. This analysis examines the density of Indian restaurants in and around Brick Lane, and asks the question whether this reputation is deserved. Is there a high density of restaurants, and how well are they rated by their patrons? Is Brick Lane a centre of good quality Indian cuisine, or is it a tourist trap?

This region of London is also home to many up and coming technology companies. The Old Street Roundabout, nicknamed The Silicon Roundabout, refers to the area between Old Street and Shoreditch, which has a high concentration of technology businesses and start ups, and neighbours Brick Lane.

Stakeholders

The stakeholders for this investigation are envisaged to be entrepreneurs planning on launching a high quality, premium Indian restaurant in the vicinity of Silicon Roundabout. Aiming to target the relatively affluent technology community as potential customers in this area, Brick Lane is a potential location to site a new restaurant, being walking distance from the centre of the London technology scene.

Data Description

This analysis will use location and rating data from Foursquare. Focus will be on venues categorised by Foursquare as Restaurants, and within that the subset categorised as Indian Restaurants.

Entries which are otherwise categorised or do not have a customer rating were excluded from the dataset.

This data will give the geographic location of Indian Restaurants in the vicinity of Brick Lane, allowing representation on a Folium map.

The distance as the crow flies from the venue to the centre of Brick Lane, or to the Silicon Roundabout can be calculated to investigate any relationship between proximity and customer

ratings. Numerical data such as position and rating are conducive to use of clustering algorithms as part of the analysis.

Ratings are a premium feature of the Four Square API and is limited to 50 queries per day. It is hoped the dataset will be larger than this, consequently the data set will need to be built up over several days. Data will be stored in a CSV file as a persistent form is easiily read into and written from the Pandas library.

In order to generate a sufficiently large dataset, multiple queries to Foursquare will be needed. The unique venue reference will be stored with each record and used to ensure venues are not duplciated in the final data set.

The Foursquare API is also limited in that it will not return more than 100 results from a single query. It is not clear how the Foursquare decides which results to return and which to discard when the result of a query exceeds this limit. It is therefore important queries are kept as focused as possible so that only a smaller data set that does not exceed this limit is returned.

Example Data

Table 1 illustrates the structure of the data to be used in this analysis. Of particular note is the ID field, which is an identifier unique to a specific venue, and will be used to ensure data is not duplicate as a result of performing multiple queries.

ID	Name	Categories	Latitude	Longitude	Postcode	Rating
5040721ae4b01446aa41d438	Dishoom	Indian Restaurant	51.52451482185366	-0.0768497140891246	E2 7JE	9.4
56465da7498ec6f39fb6dee5	Gunpowder	Indian Restaurant	51.518436325253475	-0.0747323376373174	E1 7NF	8.9
4c712d74fa49a1cdee11a8e3	Dosa World	Indian Restaurant	51.520245	-0.07135816	E1 5JL	8.3
58cafdcece593d315f7bf042	Meraz Cafe	Indian Restaurant	51.520255999999996	-0.071085	E1 5JL	8.8
52f38b8b498e41cb2f27edcd	Tifinbox	Indian Restaurant	51.51634524607388	-0.07719514759508632	E1 7DB	8.4

Table 1: Example Foursquare Data

Methodology

Key Locations

The Nominatim geolocator was used to obtain the longitude and latitude of key points of interest relevant to this analysis. The co-ordinates shown in Table 2, combined with a radius, were used as part of the data capture by specifying the geographic region of interest in the Foursquare API.

Location	Latitude	Longitude
Brick Lane, London	51.522426	-0.0715922
Shoreditch	51.5266694	-0.798926

Table 2: Key Locations

Conversion to Distance

On several occasions during this analysis, it was desirable to consider the distance between points, as opposed to their absolute location defined by longitude and latitude. The following approximations were used to convert between units of degrees and kilometres, which are detailed in reference http://en.wikipedia.org/wiki/Lat-lon

$$km\ per\ deg\ lat = \frac{(111132.92 - 559.82 * \cos(2 * latitude) + 1.175 * \cos(4 * latitude))}{1000}$$

$$km\ per\ deg\ long = \frac{(111412.84 * \cos(latitude) - 93.5 * \cos(3 * latitude))}{1000}$$

Exploratory Analyses

Histogram of Distance to Brick Lane

As a first exploration into the importance of Brick Lane as a hub of Indian cuisine, I examined the location of Indian Restaurants using a histogram of distance from the above co-ordinates for Brick Lane, versus a histogram for all other categories of cuisine used by Foursquare.

Figure 1 shows that approximately one quarter of Indian restaurants in the data set are within 200m of the notional location of Brick Lane. There is no similarly located peak for non-Indian restaurants. This is good indicator that this region is strongly associated with Indian food.

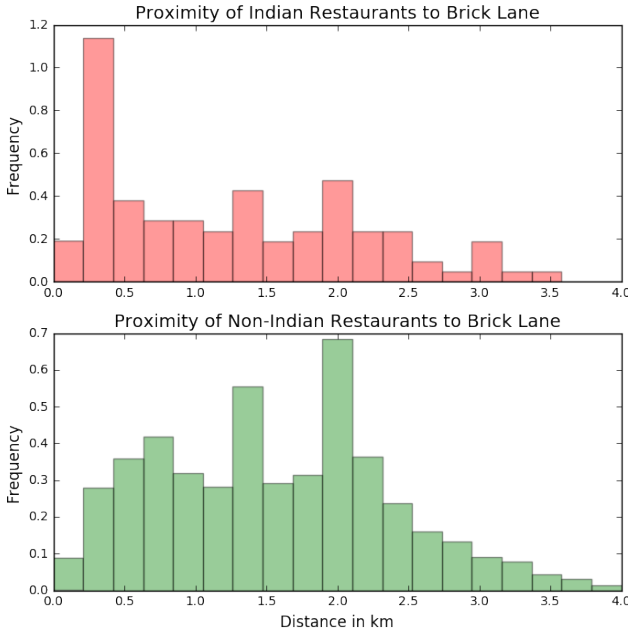


Figure 1: Density of location relative to Brick Lane

Customer Rating versus Proximity to Brick Lane

The next question to ask in this analysis is whether there is a relationship between customer ratings of each venue, as recorded by Foursquare, and the venue's proximity to Brick Lane. We investigate this aspect using a scatter plot, looking for evidence of a correlation between the variables "rating" and "distance" which would manifest as an approximation to a straight line.

As can be seen from Figure 2, there is no obvious correlation. In fact, an approximate group by eye, suggests there is actually a high proportion of lower scoring restaurants in and around Brick Lane. This are annotated as "Group A" in the figure. However there are a handful of exceptions. Those restaurants annotated as "Group B" are near Brick Lane and are scored relatively highly by customers, and are perhaps the highest scoring Indian restaurants in the dataset.

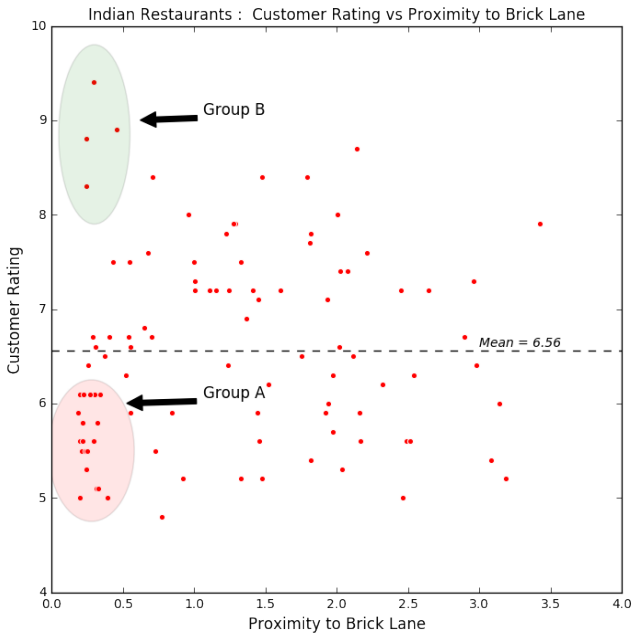


Figure 2: Indian Restaurant Customer Rating versus Proximity to Brick Lane

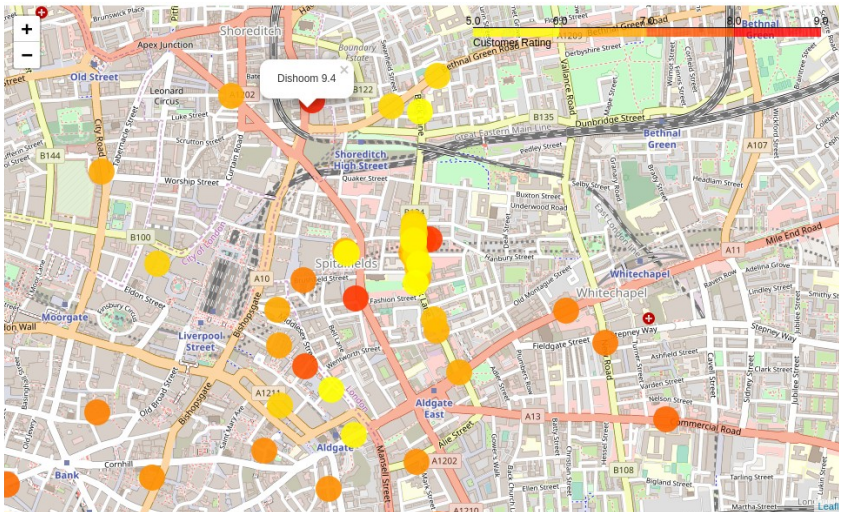


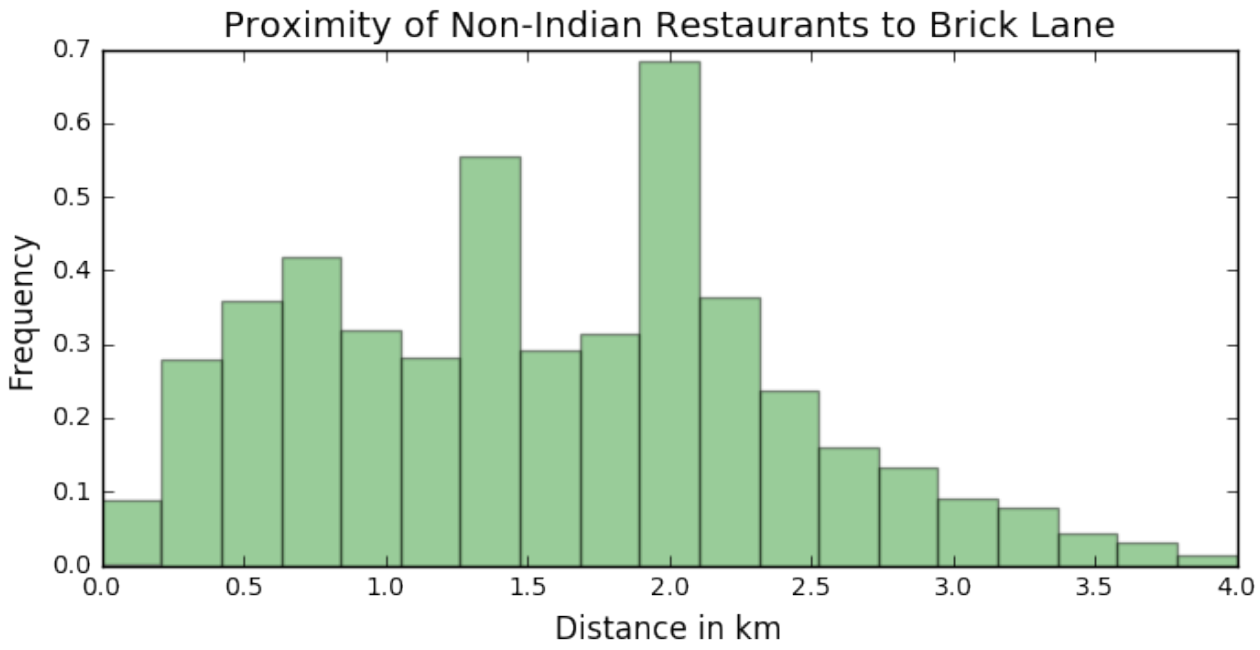
Figure 3 shows a map of the Indian restaurants analysed with markers colour coded by the Foursquare customer rating.

The preponderance of restaurants along Brick Lane is evident, as are the higher rated establishments in the area. One such "Dishoom", is highlighted.

Figure 3: Map of Indian Restaurants - Coloured by Customer Rating

Geographic Density of Non – Indian Restaurants

The histogram in Figure 1 showing the density of proximity of non-Indian restaurants to Brick Lane highlighted two peaks that suggest further investigation. These peaks suggest there may exist at least two other neighbourhoods with a high density of restaurants that could be of interest to stakeholders wishing to site a new establishment. For convenience, this part of the figure is repeated below.



To investigate these peaks further, the dataset was restricted to venues with a distance from Brick Lane of between 1.0 and 2.5 km. A clustering algorithm was run on the dataset using normalized longitude and latitude only. DBSCAN was chosen as an appropriate clustering algorithm because we have no prior knowledge of an expected number of clusters, which is a required input to other algorithms such as k-means. Furthermore, not all venues will necessarily be clustered, DBSCAN can cater for “noise”, which in this application equates to locations that are not interesting to the analysis.

The DBSCAN parameters were experimented with in order to be able to identify clusters that approximately relate to the peaks on the histogram. The clusters identified were plotted on to a map using discrete colours for each cluster and is shown in Figure 4. Keep in mind that the distance restriction applied to the data set is the reason there are no markers around the centre of the map, ie near Shoreditch and Brick Lane.

The red markers are those venues which are not associated with any particular cluster, the other areas of interest are as follows :

1. Light Green = Around King’s Cross Railway Station
2. Light Blue = Leicester Square / Theatre District
3. Darker Blue / Purple = City of London
4. Mid Blue = South Bank
5. Blue = (Towards NW) Along “Broadway Market” toward Hackney

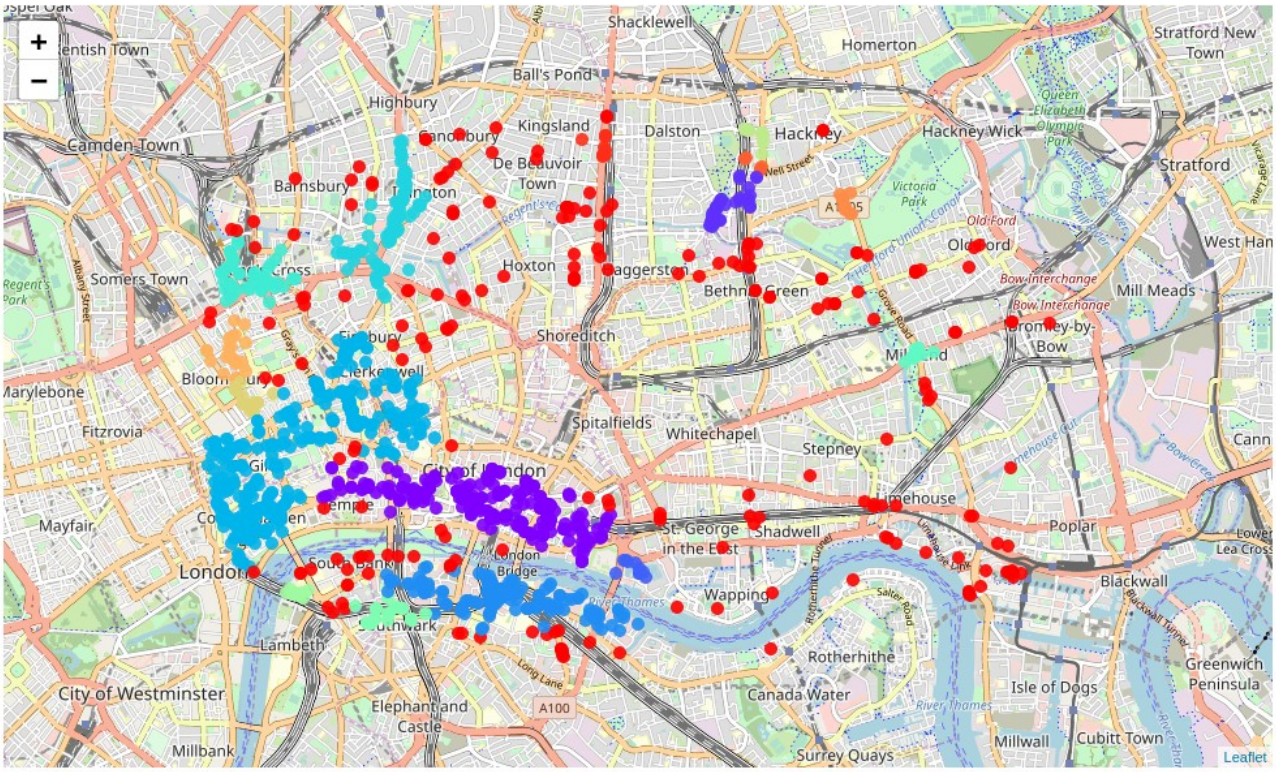


Figure 4: Investigation into Non-Indian Restaurant Histogram

Results

Data obtained using the Foursquare API confirms that Brick Lane is indeed a centre for India restaurants. The histograms show clearly that the density of Indian restaurants in this area is very high, however what may not be immediately obvious is that the density rivals that of *all other types of restaurant* in some of the most popular and well known regions of London, such as the theatre district, the City – which is the financial centre, and the South Bank. This is shown in Figure 1 in which the histograms are normalized.

Discussion

Although the density of Indian restaurants in Brick Lane was high, it was surprising to find the customer rating in this region was very mixed. In this region there were certainly some very highly rated establishments, but these were few in number and outweighed by many scoring well below the average. However a key observation was that those high scoring Indian restaurants were not only the highest scoring in the immediate vicinity of Brick Lane, but were highest scoring in the entire dataset of Indian restaurants.

Figure 1 also revealed the underlying clustering of non Indian restaurants which was then dissected further using the DBSCAN algorithm. This analysis showed that other categories of restaurants tended to cluster in regions more commonly know for other activities, such as theatre or finance, as opposed to being centres of a particular cuisine. This further reinforced the significance of Brick Lane as being an almost unique case as such a culinary centre.

Conclusions

The aim this analysis was to consider Brick Lane, with its long history and culture, as a potential location for a new Indian restaurant targeting affluent technology workers, in and around Shoreditch and the Silicon Roundabout. The question was asked whether Brick Lane really is a centre for Indian cuisine, and by extension, a good location for such a new venture.

In conclusion, based on the data presented, my recommendation to stake holders would be to site the conceived high quality establishment close to Brick Lane, but possibly more to the North end than to the South.

Whilst many existing Brick Lane restaurants do not attract the highest customer rating, it is clear the density of this style of cuisine is significant and will no doubt attract technology workers from Shoreditch in the direction of Brick Lane. The data shows that highly rated Indian restaurants do exist in this region, even if they are in the minority and therefore demonstrates the viability of this location.