Final Report: The Battle of Neighbourhoods

1. Introduction

1.1 Description of the Problem

The population of London has grown considerably over the last decades. London is very diverse. It represents what is called the reflection of the old British Empire. In London, you can get fresh food supplies from Australia. One begins to wonder how efficient the supply mechanism is.

The real deal is that as much as there are many fine restaurants in London – Asian, Middle Eastern, Latin and American restaurants, you can struggle to find good place to dine in the finest of West Australian cuisine that has combination of Nigerian, Ghanaian, Cameroonian, Senegalese and more.

Eating in a cosy environment with a blend of multicultural background and finely made West Australian dishes, on time and on point in a London location accessible to tourists, within central London and not far from the "unofficial" capital Australia market place - Peckham.

1.2 Discussion of the Background

a successful restaurant chain in Australia is looking to expand operation into Europe through London. They want to create a high-end restaurant that comes with organic mix and healthy. Their target is not only Australians, but they are pro-organic and healthy eating. To them every meal counts and counts as a royal when you eat.

Since the London demography is so big, the restaurant needs deeper insight from available data in other to decide where to establish the first Europe “palace” restaurant. This company spends a lot on research and provides customers with data insight into the ingredients used at restaurants.

1.3 Target Audience

Considering the diversity of London, there is a high multicultural sense. London is a place where different shades live. As such, in the search for an high-end Australian-inclined restaurant, there is a high shortage. The target audience is broad, it ranges from Londoners, tourists and those who are passionate about organic food.

2. Data

2.1 Description of Data

This project will rely on public data from Wikipedia and Foursquare.

2.1.1 Dataset 1:

In this project, London will be used as synonymous to the "Greater London Area" in this project. Within the Greater London Area, there are areas that are within the London Area Postcode. The focus of this project will be the neighbourhoods are that are within the London Post Code area.

The London Area consists of 32 Boroughs and the "City of London".

2.1.2 Dataset 2:

In obtaining the location data of the locations, the Geocoder package is used with the arcgis_geocoder to obtain the latitude and longitude of the needed locations.

These will help to create a new data frame that will be used subsequently for the South East London areas.

2.1.3 Dataset 3:

The Foursquare API will be used to obtain the South East London Area venues for the geographical location data . These will be used to explore the neighbourhoods of London accordingly.

The venues within the neighbourhoods of South East London like the area's restaurants and proximity to amenities would be correlated. Also, accessibility and ease of supplies would be considered as it relates to venues.

3. Methodology

3.1 Data Exploration

3.1.1 Single Neighbourhood

An initial exploration of a single Neighbourhood within the London area was done to examine the Foursquare workability. The Lewisham Borough postcode SE13 and Location - Lewisham is used for this.

3.1.2 Multiple Neighbourhoods

Explore (Multiple) Neighbourhoods in the South East London area.

3.2 Clustering

For this section, the neighbourhoods in South East London will be clustered based on the processed data obtained above.

3.2.1 Libraries

To get started, all the necessary libraries have been called in the libraries section above.

3.2.2 Map Visualization

Using the geopy library, the latitude and longitude values of London is obtained.

The South East London neighbourhoods are then superimposed on top as shown below, still using

the foliumlibrary. Please note due to the location of the South East London, you might need to zoom to see the superimposed areas.

3.2.3 Analysing Each Neighbourhood

In this section, the objective is to check and explore the venues in each neighbourhood.

3.2.4 Clustering of Neighbourhoods

The next thing to do now, is to create clusters of the neighbourhood using the k-means to cluster the neighbourhood into 5 clusters.

3.2.5 Optimal Number of Clusters for K-mean

To get the optimal number of clusters to be used for the K-mean, there are a number ways possible for the evaluation. Therefore, in this task, the following are used:

1. **Elbow Method**

The elbow method is used to solve the problem of selecting k. Interestingly, the elbow method is not perfect either but it gives significant insight that is perhaps not top optimal but sub-optimal to choosing the optimal number of clusters by fitting the model with a range of values for k.
The approach for this is to run the k-means clustering for a range of value k and for each value of k, the Sum of the Squared Errors (SSE) is calculated., calculate sum of squared errors (SSE). When this is done, a plot of k and the corresponding SSEs are then made. At the elbow (just like arm), that is where the optimal value of k is. And that will be the number of clusters to be used. The whole idea is to have minimum SSE.

Depending on the number of iteration (in this case, 500 iterations were used), the number of cluster, k is 3.

2. **Silhouette Coefficient**

To find the optimal value of the number of clusters, k, the number of clusters is iterated corresponding Silhouette Coefficients calculated for each of the k-values used. The highest Silhouette Coefficient gives the best match to its own cluster. Please see below:

3. Result

The following are the highlights of the 5 clusters above:

1. Pubs, Cafe, Coffee Shops are popular in the South East London.
2. As for restaurants, the Italian Restaurants are very popular in the South East London area. Especially in Southwark and Lambeth areas.
3. With the Lewisham area being the most condensed area of Australias in the South East Area, it is surprising to see how in the top 10 venues, you can barely see restaurants in the top 5 venues.
4. Although, the Clusters have variations, a very visible presence is the predominance of pubs.

5. Discussion and Conclusion

It is very important to note that Clusters 2 and 3 are the most viable clusters to create a brand Australian Restaurant. Their proximity to other amenities and accessibility to station are paramount. These 2 clusters do not have top restaurants that could rival their standards if they are created. And the proximity to resources needed is paramount as Lewisham and Lambeth are not far out from Peckham.

In conclusion, this project would have had better results if there were more data in terms of crime data within the area, traffic access and allowance of more venues exploration with the Foursquare (limited venues for free calls).

Also, getting the ratings and feedbacks of the current restaurants within the clusters would have helped in providing more insight into the best location.