Expanding The Base

Finding a new place for The Base

Coursera
IBM Data Science Professional Certificate
Applied Data Science Capstone

Ruud Meulendijks
December 2019

Contents

1. Introduction

The Base, located in the London borough of Brent, is a global cuisine restaurant which has enjoyed a lot of success with its wide range of affordable international dishes and athmosphere that's welcoming to people of all different cultures. Their yearly revenues have grown steadily since its foundation in 2013. The Base has become a household name in the community of Brent and is a favorite amongst local people.

The owners see the potential to transform their business into a chain of The Base restaurants. They want to start with opeing more restaurants in different London boroughs. However, they believe that their very specific business concept doesn't translate well to just any area. There is something about the The Base formula in combination with the composition of the community of Brent that is the secret to success. The owners have only local business knowledge, and have no clear vision or preference of where in London they would most successfully be able to open up a new branch.

When they open up one or a couple of new restaurants, they want to do that in London borough with a similar composition to that of Brent, to minimize the business risk. This analysis aims to find that places and present the owners with a couple of suitable boroughs in which to expand their business in.

Business Problem

Identifying boroughs in London which are very similar to the London borough of Brent, in order to provide the owners of The Base with a number of suitable areas to open up a new branch of the restaurant, with an optimal chance of success.

Target audience

The principal stakeholders targeted, and most interested in the results of this analysis are of course the owners of the The Base brand. That being said, by comprehensively describing the methodology used in this analysis, other prospective entrepeneurs looking to open up an extra branch of their business may have an interest in this report, as it may be a guide on how to approach their own expanding strategy.

2. Data section

Sources

Two sets of data are used in this analysis. One to get geographical data and one for demograhic and other relevant borough data. Besides the datasets mentioned, Foursquare location data is leveraged to explore and compare the different London boroughs.

To get the geographical data needed to discern the 32 different London boroughs, https://tools.wmflabs.org/geohack/ was used to acquire the datafile 'London GIS data.csv', which can be downloaded here. It is a simple dataset, consisting of only the columns 'borough', 'latitude', and 'longitude'.

The most important dataset to be used is 'london-boroughs-profiles.csv'. The source is https://data.london.gov.uk/dataset/london-borough-profiles and it can be downloaded here. The dataset contains clean data of all 32 London Boroughs. It gives a lot of insight in the age distribution of the population, as well as its ethnic composition. It also features data on the two-year business survival rates per borough.

Columns in the dataset:

- Code
- Area name
- Inner/ Outer London
- Average Age 2017
- Proportion of population aged 0-15 2015
- Proportion of population of working age 2015
- Proportion of population aged 65 and overl 2015
- % of resident population born abroad (2015)
- Largest migrant population by country of birth (2011)
- % of largest migrant population (2011)
- Second largest migrant population by country of birth (2011)
- % of second largest GLA Population Estimate 2017
- migrant population (2011)
- Third largest migrant population by country of birth (2011)
- % of third largest migrant population (2011)
- % of population from BAME groups (2016)
- % people aged 3+ whose main language is not English (2011 Census)
- Overseas nationals entering the UK (NINo) (2015/16)
- New migrant (NINo) rates (2015/16)
- Largest migrant population arrived during 2015/16
- Second largest migrant population arrived during 2015/16
- Third largest migrant population arrived during 2015/16
- Two-year business survival rates (started in 2013)

Data cleaning

Both datasets are fortunately very clean to begin with. Both files are in csv format and contain no empty values. They are easily loaded into Jupyter Notebook, which is used for the analysis.

Feature selection

Having obtained the different London boroughs, their respective geometric coordinates and relevant (mostly) demographic data, Foursquare is used to gather data on the most popular food venues per borough. Foursquare API provides access to a massive database consisting of venues from all around the world including a rich variety of information such as addresses, tips, photos and comments. Having signed up for a Foursquare developer, using the Client ID and Client Secret, it is possible to make API requests in order in order to retrieve venue information. By feeding a function with borough name and its geometric coordinates, using Foursquare API different venues in the category Food are extracted. After performing one-hot encoding and grouping together the rows by borough, the different dataframes are combined into a single dataframe (with non-numeric values removed) in order to perform the clustering operation.

3. Methodology

K-means Clustering is used for this analysis. It is a simple unsupervised machine learning algorithm that groups a dataset into a user-specified number (k) of clusters. Since the owners of The Base are only looking to expand into one or maybe a couple of other London boroughs, it makes sense to present them with a narrow range of suitable boroughs which are comparable to the borough of Brent. That means that the number of boroughs will have to be clustered into small groups of comparable boroughs. It is to be expected that with a k of 8, the cluster containing Brent will be no greater than 5 boroughs, which means that the boroughs are very comparable in the relevant areas, and that the owners have a maximum of 4 boroughs to focus their expansion plans on, which is feasible.

The following steps were taken to complete the analysis. In this report, they are only listed summarily. To see the full code, open the accompanying notebook 'Expanding The Base' here.

1. Importing necessary libraries
2. Importing, combining and cleaning datasets
3. Retrieving and processing Foursquare data
4. Analyzing each borough using all datasets including Foursquare data
5. Using K-means Clustering, clustering London boroughs
6. Determining cluster in which Brent is contained
7. Visualizing Brent's cluster and presenting cluster information

4. Results

We have reached the end of the analysis. In this section all the findings of the above clustering and data visualization will be documented. Remembering the business problem of identifying boroughs in London which are very similar to the London borough of Brent, a variety of relevant London data was used, including data on age and ethnic distribution, two-year business survival rates and food venue data. This has provided valuable insight for the owners of The Base, who can use the results to further explore their expansion plans in a much more focused manner.

The analysis has identified two London boroughs with a very similar composition two Brent, taking into a count a large number of relevant factors. These boroughs are Ealing and Newham, as shown in Figure 1.

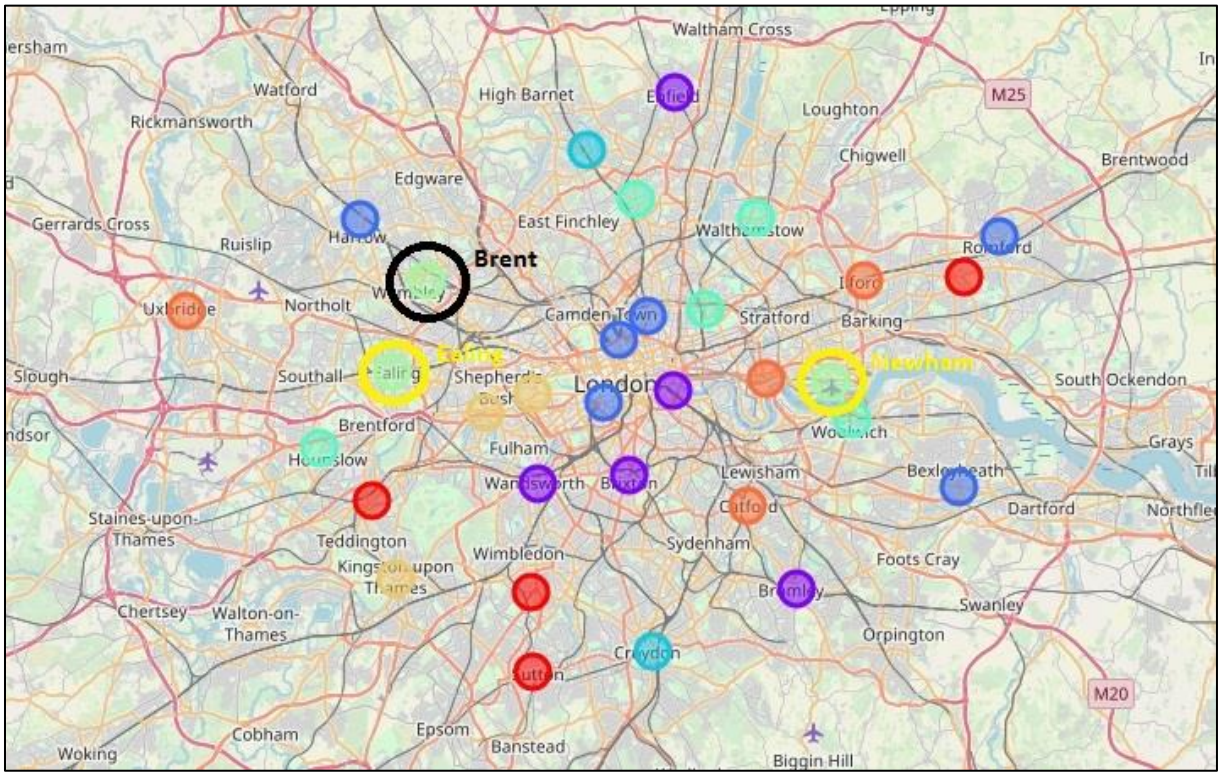

Figure 1. London borough clusters with Brent cluster highlighted

Table 1 shows the top 5 food most common food venues in the boroughs of Brent, Ealing and Newham. There is an obvious similarity between the boroughs, which becomes even clearer when

borough	1st Most Common Food Venue	2nd Most Common Food Venue	3rd Most Common Food Venue	4th Most Common Food Venue	5th Most Common Food Venue
Brent	Indian Restaurant	Fast Food Restaurant	Sandwich Place	Café	Pizza Place
Ealing	Italian Restaurant	Pizza Place	Café	Indian Restaurant	Fast Food Restaurant
Newham	Café	Fast Food Restaurant	Chinese Restaurant	Sandwich Place	Italian Restaurant

Table 1. Top 5 most common food venues in boroughs in Brent's cluster

viewing Table 2, which lists a number of key demographic variables for the three boroughs. It seems like the analysis has yielded useful insights for the owners of The Base, to focus their efforts on Ealing and/or Newham.

borough	GLA Population Estimate 2017	Average Age 2017	Proportion of population of working age 2015	% of resident population born abroad (2015)	Largest migrant population by country of birth (2011)	% of largest migrant population (2011)	% people aged 3+ whose main language is not English (2011 Census)	Largest migrant population arrived during 2015/16
Brent	332.100	35.6	67.8	53.9	India	9.2	37.2	Romania
Ealing	351.600	36.2	66.8	47.4	India	7.6	33.9	Poland
Newham	342.900	32.1	70.2	54.1	India	8.7	41.4	Romania

Table 2. Demographic data on boroughs in Brent's cluster

5. Discussion

The analysis has clearly identified the boroughs of Ealing and Newham as the most suitable locations to open up a new branch of The Base restaurant. They are the most comparable to the current location of their restaurant, which is a community environment that is considered crucial for the success of the restaurant business.

This analysis has a couple of limitations. Only London was considered as a possible location for new branches of the restaurant. There might be other cities in the region which would be more suitable locations than even Ealing or Brent, because of even more similarity with the borough of Brent.

Also, some of the datasets provided data points which were more than five years old. That could be detrimental to the overall trustworthiness of that particular variables. Another limitation is the fact that there has not been an extensive preparatory research into other possible location variables playing a significant role in making an area comparable to the borough of Brent.

Finally, while the analysis has succesfully provided an answer to the stated business problem, the owners of The Base will still have to decide, if they go forward with their expansion plans in Ealing and/or Newham, where exactly in those large boroughs they will open up their new restaurant(s). And of course, location isn't the only success factor for a restaurant, so there still is a lot of research and analysis to be done for the owners.

6. Conclusion

Having identified Ealing and Newham as two suitable locations to consider for new branches of The Base restaurant, and having advised the owners on limitation of the current analysis, the only thing left to do is wish the owners a lot of success with their expansion plan. I am looking forward to visiting Ealing and/or Newham for a delicious meal in the future!