Best location to open a café in London

Introduction

Background:

I am assuming the position of a business man who wants to raise funding from stakeholders to open a café in London. To ensure I have the best chances of succeeding I would like to open a cafe in an area where cafe's are one of the least (or at least less) cafe dense areas. I will also like to open a café in one of the areas that are most populated. The logic is that a region that has fewer coffee shops would be less competitive and a larger population would mean a larger potential customer base.

Data description:

I have sourced data from Wikipedia which provided borough's in London as well as the population in the respective location as of September 2013. I have also used Foursquare's API to pull data on popular venues in each borough.

Methodology:

My intended technique to use was clustering so that I can report to the stakeholders a number of areas in London that are best suited for opening a café. This was done by grouping each borough with the number of café's they have as well as the population of the respective borough. I then scaled the data to avoid population dominating number of café's as a feature due to its considerably larger magnitude. I then hoped to have a cluster of borough's that represented regions with low number of cafes and a large population. This is the most profitable regions for myself to open a café.

Data exploration consisted of mainly looking at the data using .head() frequently to ensure my data was in the form that I needed it to be.

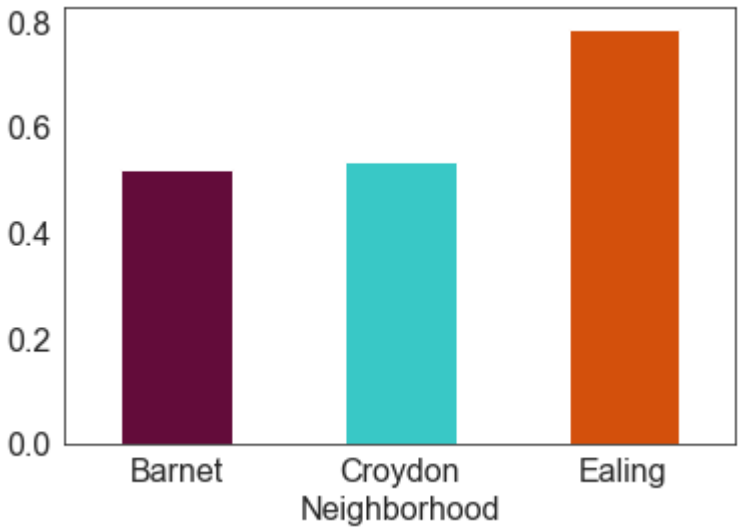
	0	1	2	3	4	5	6	7
0	Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est) [1]
1	Barking and Dagenham [note 1]	NaN	NaN	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194352
2	Barnet	NaN	NaN	Barnet London Borough Council	Conservative	North London Business Park, Oakleigh Road South	33.49	369088
3	Bexley	NaN	NaN	Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236687

This was the unclean version of the dataset I needed. Ultimately, I needed the dataset to be in the following format: Columns: Borough, Population, Longitude and Latitude. After cleaning the dataset, I managed to get it in the following format:

	Borough	Population (2013 est)[1]	Co-ordinates	Longitude	Latitude
1	Barking and Dagenham	194352	[51.3339, 0.0921]	0.0921	51.3339
2	Barnet	369088	[51.3731, 0.0906]	0.0906	51.3731
3	Bexley	236687	[51.2718, 0.0902]	0.0902	51.2718
4	Brent	317264	[51.3332, 0.1654]	0.1654	51.3332
5	Bromley	317899	[51.2414, 0.0111]	0.0111	51.2414

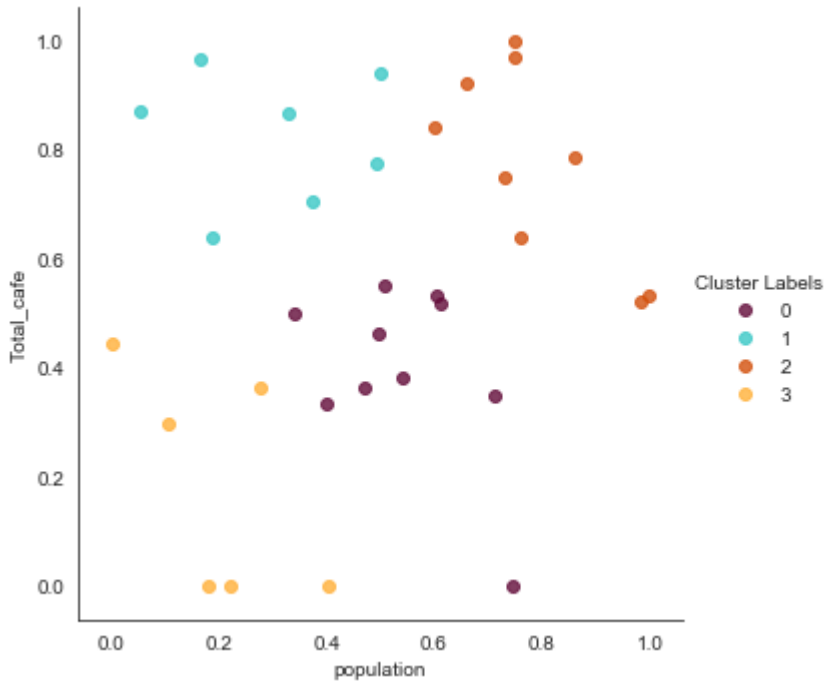
For me to be able to perform clustering analysis I needed a dataframe with the count of the number of cafes in respective region (sourced from Foursquare) and then group it by the borough. This will be appended to the dataframe with population in it. I can then scale the population and count of café shops to ensure non biased estimates are formed. The following dataframe is produced:

	Cluster Labels	Neighborhood	population	Longitude	Latitude	Total_cafe
0	3	Barking and Dagenham	0.178478	0.0921	51.3339	0.000000
1	2	Barnet	0.983127	0.0906	51.3731	0.519481
2	1	Bexley	0.373429	0.0902	51.2718	0.705882
3	0	Brent	0.744481	0.1654	51.3332	0.000000
4	2	Bromley	0.747405	0.0111	51.2414	0.969697
5	0	Camden	0.341341	0.0732	51.3144	0.500000
6	2	Croydon	1.000000	0.0552	51.2217	0.533333
7	2	Ealing	0.860664	0.1832	51.3047	0.786885
8	2	Enfield	0.759493	0.0448	51.3914	0.640000
9	1	Greenwich	0.499240	0.0353	51.2921	0.941176



This visualisation represents region with greater than 80 percent of the max population (>298201) with the mean café amount on the y axis. Ealing has a high number of café's whereas Barnet has comparatively fewer making it a more feasible location of opening a café.

Once clustering is applied, I can visualise the clusters to see which clusters have my intended cluster values (low number of café's and large population). This was the output of visualising the clusters:



Results:

As you can see the cluster that best fits my criteria is cluster label 0. I then retrieved all boroughs that were in cluster '0'.:

Out[76]:

	Cluster Labels	Neighborhood	population	Longitude	Latitude	Total_cafe
0	0	Brent	0.744481	0.1654	51.3332	0.000000
1	0	Camden	0.341341	0.0732	51.3144	0.500000
2	0	Hackney	0.468714	0.0319	51.3242	0.363636
3	0	Haringey	0.496376	0.0643	51.3600	0.463768
4	0	Havering	0.398263	0.1101	51.3452	0.333333
5	0	Hillingdon	0.604224	0.2834	51.3239	0.533333
6	0	Redbridge	0.610974	0.0427	51.3332	0.516129
7	0	Tower Hamlets	0.540141	0.0021	51.3036	0.380952
8	0	Waltham Forest	0.507478	0.0048	51.3527	0.551724
9	0	Wandsworth	0.713407	0.1128	51.2724	0.347826

These are the boroughs that are best suited to open a café based on the fact that there are low number of café's but large population.

Discussion:

From the results discussed above; there are a few observations that struck me. One observation is that the Borough of Brent has 0 reported café's or virtually 0 compared to other boroughs. This was a bit surprising but from further research it seems that four square doesn't have much data on the café's open in Brent. I should then exclude this as being an accurate, representative result. Other boroughs that seem promising include Tower Hamlets and Wandsworth. I would bring these results to the stakeholders and suggest opening a café in Wandsworth as the price of opening a business there is significantly cheaper compared to Tower Hamlets.

Conclusion:

Based on the above results I was able to successfully find a group of areas which best suit a café opening; the next steps would be to discuss with stakeholders the cost of opening a café in each of these areas to make the final decision on where to open the café.