Coursera Capstone Project

By Sana Syed

The Battle of the Neighborhoods

Part of the IBM Data Science Professional Certificate

Introduction / Business Problem

Chai San wants to open his first ever Thai Restaurant in England but, having never visited the country before, has no idea as to where would be the best location to begin looking in. We are tasked with identifying the best location for his new business venture.

The goal put-forward to us is to identify a location which houses zero other Thai restaurants in close proximity. The location should also not already be overcrowded with restaurants of any type and should be in an area with a high Thai population.

In order to identify which location would suit Chai San's restaurant the best, we will gather and analyze publicly available data of residents of England and facilities for each district. This will ensure that the location selected is optimal for the target demographic with a high density of potential customers and meets all of the goals set-forth.

Data

Based on the given business problem, we can determine that we will require the following data to perform our analysis:

- Population data of ethnic groups across England
- List of existing restaurants in the neighborhood
- List of existing Thai restaurants in the neighborhood
- List of nearby venues and their types in the neighborhood
- List of the boroughs in the chosen county

The following sources will be used to gather the data required:

• Foursquare API for restaurant/venue information

• Office of National Statistics for Ethnic Populating Data

• http://martinjc.github.io For GeoJSON Data of England

• Wikipedia for a list of London boroughs and lat/lon values

Methodology

The aim of our focus was to identify areas of England which harbour the perfect conditions for our client to open his new Thai restaurant. The perfect conditions, as set out in the introduction, were areas that contained a high Asian population, had a relatively low restaurant density and were distant from other nearby Thai restaurants.

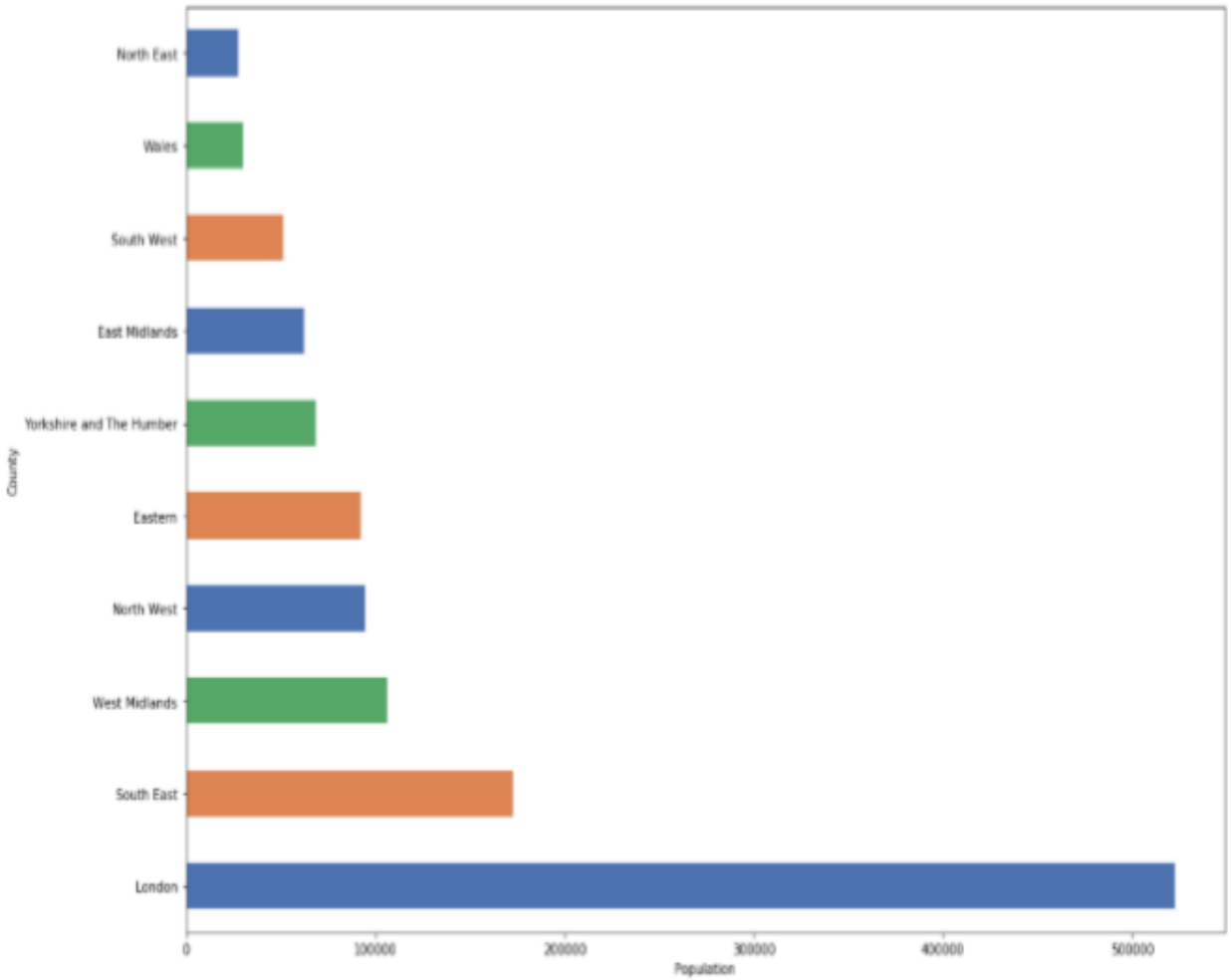
During the data collection phase of this project, we gathered Population data from the Office of National Statistics to identify where the Asian communities in England are mostly located. As the dataset does not explicitly list the Thai population in England, the decision was made to re-classify our target demographic as “Chinese” and “Other Asian, as these two groups are considered most likely to be of Thai decent.

Figure

Figure 1 Population Data (England)

	Ethnic Group	North East	North West	Yorkshire and The Humber	East Midlands	West Midlands	East	London	South East	South West	Wales
0	White: English/Welsh/Scottish/Northern Irish/B...	2431423	6345089	4531137	3871146	4434333	4986170	3668284	7358998	4855675	2855450
1	White: Irish	8035	64930	26410	28676	55216	55573	175974	73571	28616	14086
2	White: Gypsy or Irish Traveller	1684	4147	4378	3418	4704	8305	8186	14542	5631	2785
3	White: Other White	34425	155570	130031	143116	139386	240286	1033981	380709	156506	55932
4	Mixed/multiple ethnic groups: White and Black ...	5938	39204	16241	43404	68533	37222	119425	41980	25669	11099
5	Mixed/multiple ethnic groups: White and Black ...	3549	18392	9321	8814	9232	15388	65478	22825	8550	4424
6	Mixed/multiple ethnic groups: White and Asian	8022	30529	26008	23688	32561	32226	101300	38764	21410	9019
7	Mixed/multiple ethnic groups: Other Mixed	4940	22766	15988	15318	21388	27180	118875	40395	18255	8979
8	Asian/Asian British: Indian	13817	107353	69252	168928	218439	86736	542857	152132	34388	17256
9	Asian/Asian British: Pakistani	19831	189435	125892	68940	217348	66270	223797	99240	13922	12229
10	Asian/Asian British: Bangladeshi	10972	45897	22424	15258	52477	32992	222537	27951	8416	10687
11	Asian/Asian British: Chinese	14284	48049	28435	24404	31274	33501	124250	53061	22243	13618
12	Asian/Asian British: Other Asian	13895	49750	39981	37893	74997	58871	398535	119952	29068	16318
13	Black/African/Caribbean/Black British: African	10782	59178	46033	41768	64253	69925	573931	87345	24226	11887
14	Black/African/Caribbean/Black British: Caribbean	1193	23131	23430	28913	86794	33614	344597	34225	15329	5809
15	Black/African/Caribbean/Black British: Other B...	1045	15460	10892	10801	15078	13903	170152	14443	10121	2580
16	Other ethnic group: Arab	5850	24528	21340	9796	18079	30367	186020	35363	5092	9635
17	Other ethnic group: Any other ethnic group	5205	19688	19570	15989	15825	18474	175023	31748	9917	5663

Figure 2 Asian Population (England)



Using this information, we were able to direct our focus to London, which had a far bigger Asian (Chinese or Other) population than anywhere else in the country with approximately 523,000 citizens. This then allowed us to propagate a list of 32 boroughs within the London area for which we could analyse further in order to determine an exact location for our clients restaurant.

Figure 3 Boroughs of London

	Borough	Latitude	Longitud
0	Barking and Dagenham	51.5607	0.1557
1	Barnet	51.6252	-0.1517
2	Bexley	51.4549	0.1505
3	Brent	51.5588	-0.2817
4	Bromley	51.4039	0.0198
5	Camden	51.529	-0.1255
6	Croydon	51.3714	-0.0977
7	Ealing	51.513	-0.3089
8	Enfield	51.6538	-0.0799
9	Greenwich	51.4892	0.0648
10	Hackney	51.545	-0.0553
11	Hammersmith and Fulham	51.4927	-0.2339
12	Haringey	51.6	-0.1119
13	Harrow	51.5898	-0.3346
14	Havering	51.5812	0.1837
15	Hillingdon	51.5441	-0.476
16	Hounslow	51.4746	-0.368
17	Islington	51.5416	-0.1022
18	Kensington and Chelsea	51.502	-0.1947
19	Kingston upon Thames	51.4085	-0.3064
20	Lambeth	51.4607	-0.1163
21	Lewisham	51.4452	-0.0209
22	Merton	51.4014	-0.1958
23	Newham	51.5077	0.0469
24	Redbridge	51.559	0.0741
25	Richmond upon Thames	51.4479	-0.326
26	Southwark	51.5035	-0.0804
27	Sutton	51.3618	-0.1945
28	Tower Hamlets	51.5099	-0.0059
29	Waltham Forest	51.5908	-0.0134
30	Wandsworth	51.4567	-0.191
31	Westminster	51.4973	-0.1372

For each of the 32 boroughs, by querying the Foursquare API, we identified and gathered a list of locations for all 1,477 venues in London and, more specifically, the locations of all Thai and non-Thai restaurants in London too. We found that there was a total of 348 restaurants in London, 16 of which are Thai restaurants.

This data would prove vital in allowing us to pinpoint the location of our client's new restaurant as we begin our analysis. The below table details the venues and locations for one borough, Barking and Dagenham.

Figure 4 Barking and Dagenham Venues

	Venue Name	Cetegory	Lat	Long
0	Central Park	Park	51.55956	0.161981
1	Crowlands Heath Golf Course	Golf Course	51.562457	0.155818
2	Robert Clack Leisure Centre	Martial Arts Dojo	51.560808	0.152704
3	Beacontree Heath Leisure Centre	Gym / Fitness Center	51.560997	0.148932
4	Morrisons Becontree Heath	Supermarket	51.559774	0.148752
5	Becontree Heath Bus Station	Bus Station	51.561065	0.150998
6	Dagenham Swimming Pool	Pool	51.560946	0.150054

Once in possession of venue data from the Foursquare API, further analysis could be conducted on each individual borough. We identified which boroughs contained the highest number of restaurants within a 500-meter radius and their locations. We also identified the locations of all Thai restaurants within each borough too as these were to be avoided at all cost.

Figure 5 Total Restaurants by Borough

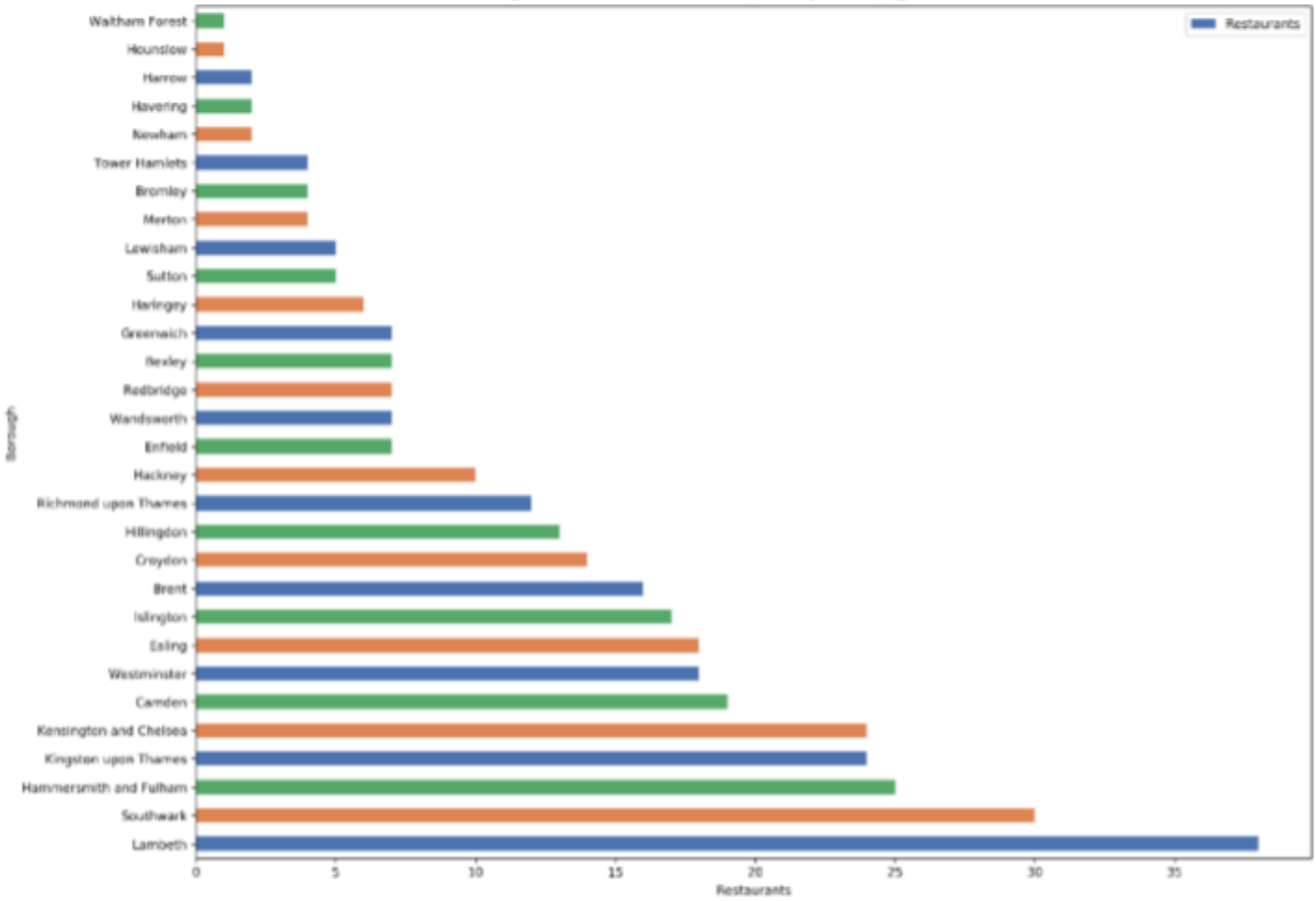


Figure 6 Thai Restaurants by Borough

	Borough	Borough Lat	Borough Lon	Venue	Venue Latitude	Venue Longitud	Venue Category
1	Camden	51.529	-0.1255	Supawan	51.532009	-0.120637	Thai Restaurant
2	Croydon	51.3714	-0.0977	Thai Express	51.375029	-0.093638	Thai Restaurant
3	Ealing	51.513	-0.3089	@ Thai Canteen	51.513037	-0.306322	Thai Restaurant
4	Greenwich	51.4892	0.0648	Tai Tip Mein Noodle Bar	51.490636	0.06953	Thai Restaurant
5	Hammersmith and Fulham	51.4927	-0.2339	Poppy's	51.494477	-0.231942	Thai Restaurant
6	Hammersmith and Fulham	51.4927	-0.2339	The Hammersmith Cafe	51.493051	-0.234285	Thai Restaurant
7	Islington	51.5416	-0.1022	Isarn	51.538505	-0.102436	Thai Restaurant
8	Islington	51.5416	-0.1022	Rosa's Thai Cafe	51.537563	-0.103135	Thai Restaurant
9	Kingston upon Thames	51.4085	-0.3064	Busaba Eathai	51.410646	-0.308153	Thai Restaurant
10	Lambeth	51.4607	-0.1163	Kaosarn	51.462152	-0.111493	Thai Restaurant
11	Lambeth	51.4607	-0.1163	YUM-D	51.461797	-0.113275	Thai Restaurant
12	Southwark	51.5035	-0.0804	Kin+Deum	51.502471	-0.082046	Thai Restaurant
13	Southwark	51.5035	-0.0804	Suchard Thai Restaurant	51.504545	-0.08254	Thai Restaurant
14	Wandsworth	51.4567	-0.191	The Thai	51.459859	-0.189345	Thai Restaurant

These locations were then superimposed onto a map of London for easy visualization of the surrounding area to aid with selecting the best location. This map allowed us to clearly visualize the distance between each borough, nearby venues, the city center and more.

Figure 7 London Restaurants (All)

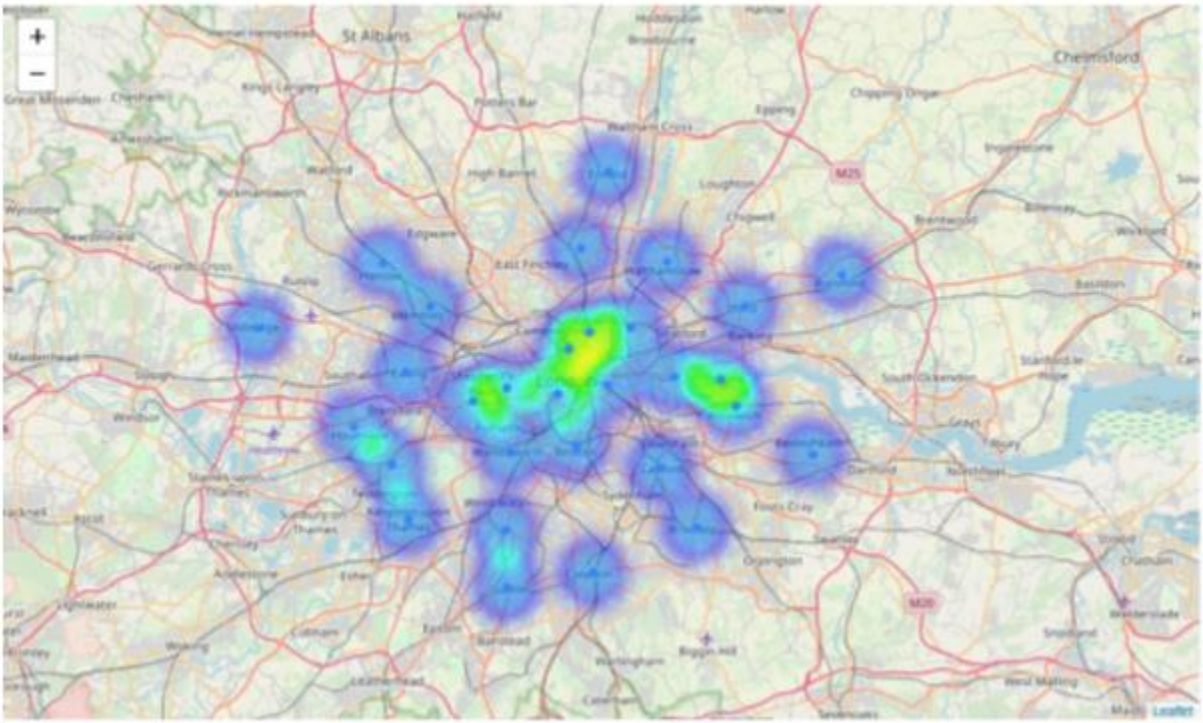
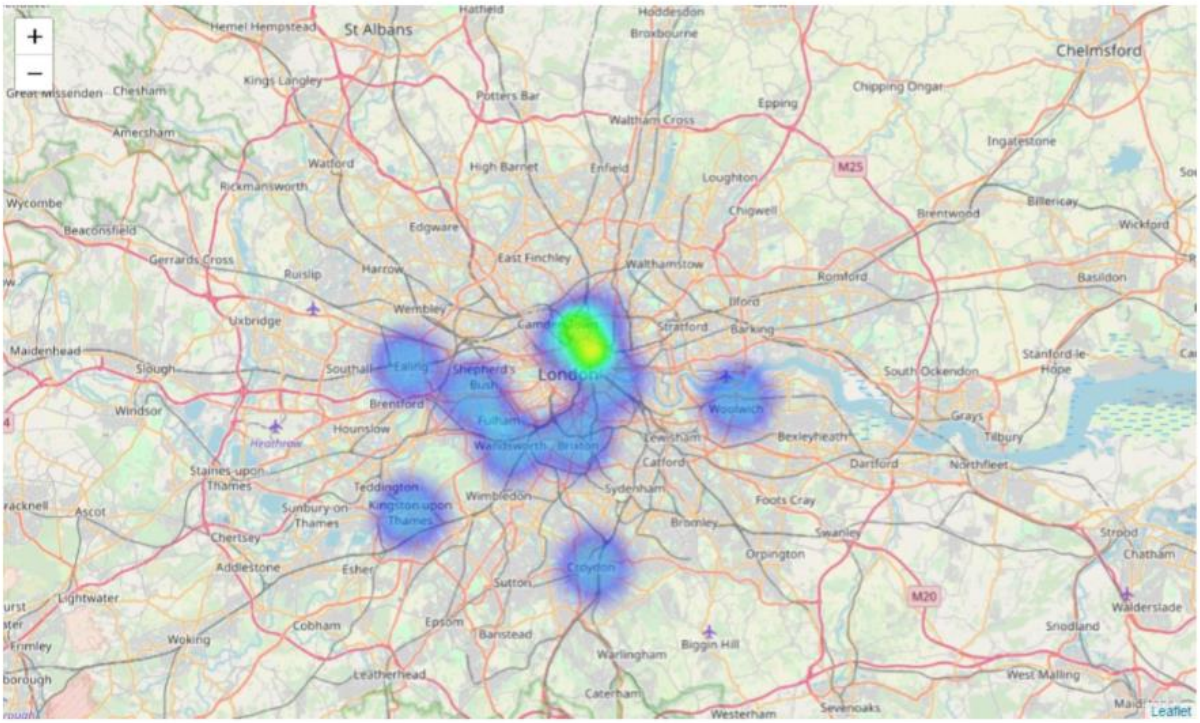


Figure 8 London Restaurants (Thai)



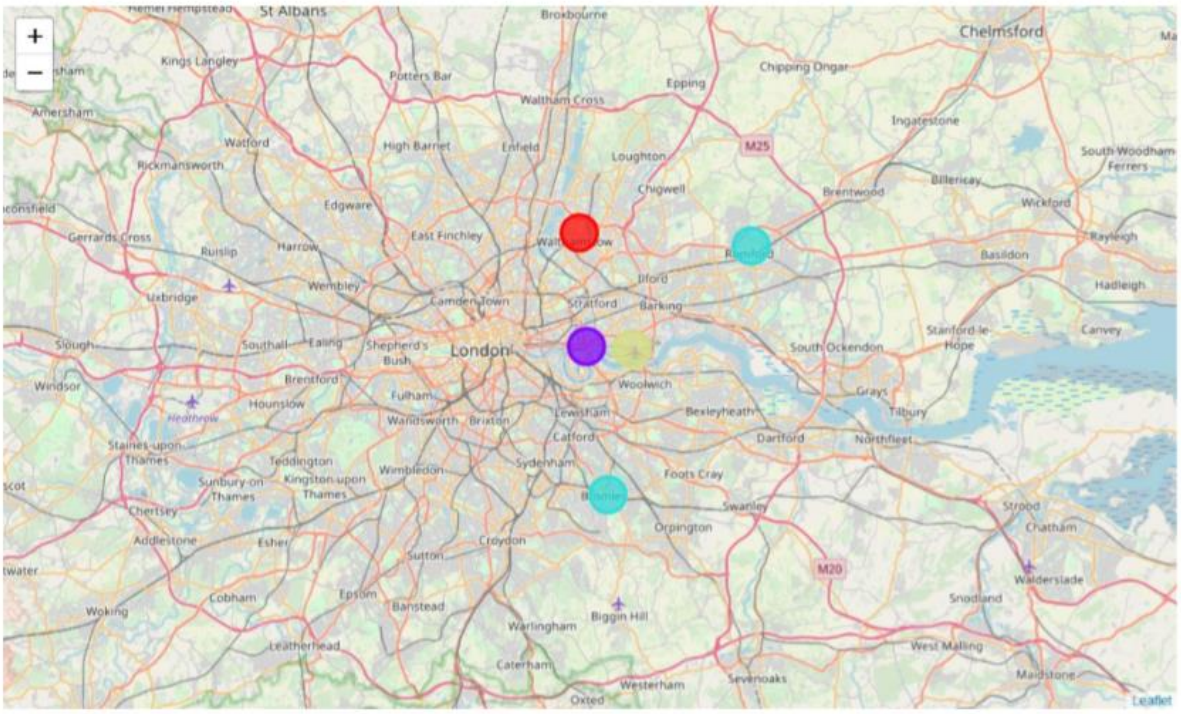
After visualizing the venues on a map, all boroughs which both contained a Thai restaurant and were densely populated with restaurants of any type were removed from the dataset. This left us we a dataset containing only 5 locations which fit our client's goals, each in prime locations which ticked all three of the boxes our client asked for.

Machine Learning techniques could then be used on the remaining boroughs in the dataset to cluster boroughs based on their similarity. From this we can identify which boroughs were close to hotels, pubs, Airports and other venues that would be a good location for our clients' new restaurant in terms of foot traffic.

Figure 9 Clustered Boroughs

Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Waltham Forest	0	Grocery Store	Pub	Gym	Beer Store	Gym / Fitness Center
Tower Hamlets	1	Hotel	Italian Restaurant	Coffee Shop	Pizza Place	Sandwich Place
Bromley	2	Coffee Shop	Clothing Store	Burger Joint	Pizza Place	Gym / Fitness Center
Havering	2	Clothing Store	Coffee Shop	Shopping Mall	Café	Fast Food Restaurant
Newham	3	Hotel	Airport	Duty-free Shop	Chinese Restaurant	Light Rail Station

Figure 10 Clustered Boroughs



Results and Discussion

Our study managed to accurately find and select a series of locations which meet the business problem given by our client. 5 boroughs which had a high Asian population, the lowest restaurant density and also contained zero Thai restaurants were identified and probed.

We discovered that the boroughs of both Bromley and Havering were very similar because they both have Clothing Stores and Coffee Shops as their first and second most popular venues respectively. One downside they both share however is that they are both located the furthest distance away from the city center which could mean less foot traffic compared to other locations.

Newham was found to be close to both a Hotel and an Airport, which would be a great location for attracting many potential customers with potentially a very high traffic stream passing through. Waltham Forest was found to be close to a Pub, Gym and Social Centre, which also could be a high traffic area for a proposed business.

Conclusion In this study we analysed the ideal location for a new Thai restaurant based on the following criteria: The chosen location must be in an area of high Asian population, be distant from any other Thai restaurants and also not be in an area which contains a high density of restaurants of any type.

CONCLUSION

Population data of ethnic groups within the UK accurately identified London as the county with the highest Asian population and therefore London was selected for further probing. Within London, each borough was analyzed based on the number of venues, number of restaurants and number of Thai restaurants that they each contained.

These boroughs were then sorted and filtered based on the criteria given to leave just 5 boroughs which could be labelled as prime locations for the new restaurant. Machine Learning was applied to the 5 selected boroughs to cluster each borough based on similarity and their 5 most common venues.

These 5 boroughs could now be put-forward to the client for his consideration and, if the client so wishes, each borough could be probed further using the same techniques on a more concentrated level.