Battle of Neighborhoods

CAPSTONE FINAL PROJECT

1. Problem description

- The project goal is to use analytical approach to recomend the best location for a Chinese restaurant in London, England. It will be used machine learning techniques, data analyses and, data visualization.

2. Data presentation

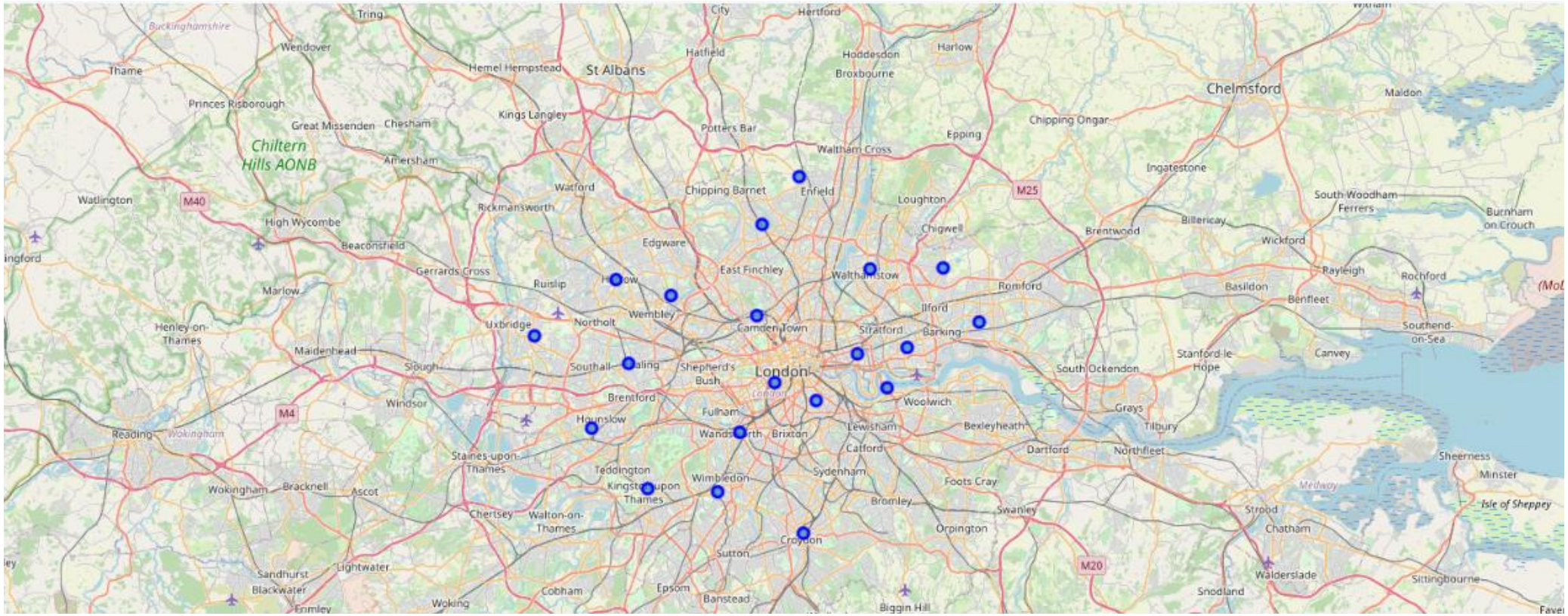
- The data source is from two sites:
- The Foursquare Api: It will be used to obtain the most common venues per neighborhood in London. It will allow the user to know how the city's venues are distributed, what are the most common places for leisure, and in general, it will provide an idea of what people's likes are.
- Wikipedia's Ethnic groups in London webpage: This site provides information about ethnicity of population in London. The webpage is scraped using BeautifulSoup4, and the table containing Asian population of London is converted into DataFrame. The data contains information about the inmigrant population per borough and per nationality. This data will be analyzed in such a way that one could determine the best location of r anew venue/restaurant/other based on people's nationalities. For the sake of simplicity, it will be assumed for this exercise that people's likes varies according to their nationality, and that people from one specific country will be more attracted to place that matches the environment and culture of their own countries, rather than the ones from foreign countries.
- The data is acessed in the link:https://en.wikipedia.org/wiki/Ethnic_groups_in_London

Exploring Asian population in London

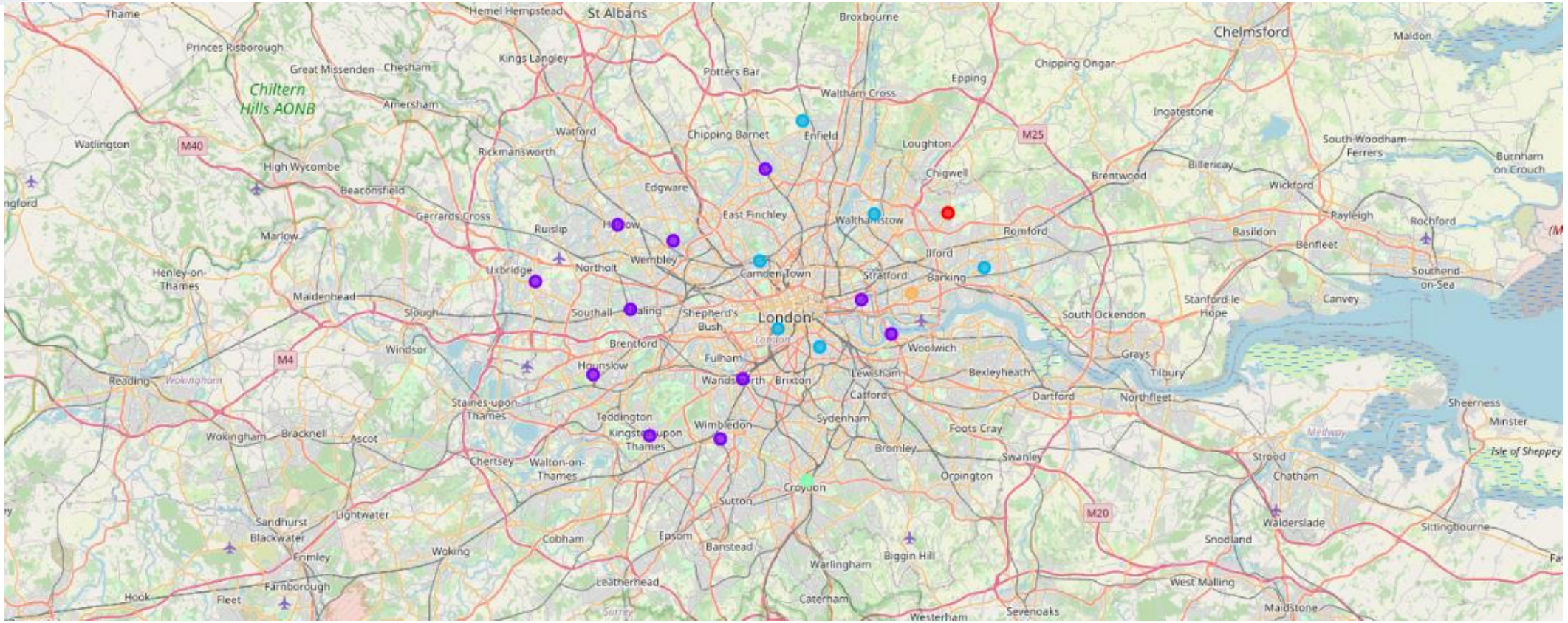
	Indian Population	Pakistani Population	Bangladeshi Population	Chinese Population	Other Asian Population	Total Asian Population
count	20.000000	20.000000	20.000000	20.000000	20.000000	20.00000
mean	24000.850000	10241.000000	9803.400000	4188.550000	15970.700000	64204.50000
std	20240.083397	9269.957588	18793.005256	2098.051213	7577.190808	35563.84754
min	5819.000000	1489.000000	892.000000	1315.000000	5135.000000	26152.00000
25%	7380.250000	2594.000000	1776.750000	2626.250000	10021.250000	32969.00000
50%	10391.000000	7902.000000	2604.500000	3482.500000	14454.500000	57008.00000
75%	43278.000000	13852.250000	6649.500000	5275.000000	20792.250000	100781.25000
max	63051.000000	31051.000000	81377.000000	8259.000000	31570.000000	133895.00000

- As shown above the mean Chinese Population is around 4200 and the maximum population in a neighborhood is 8259.

London's map to see the neighborhoods distribution



Visualize our Clusters distribution on London's Map



Cluster 1

	London Borough	Indian Population	Pakistani Population	Bangladeshi Population	Chinese Population	Other Asian Population	Total Asian Population	Neighborhood
1	Redbridge	45660	31051	16011	3000	20781	116503	Redbridge

Cluster 2

	London Borough	Indian Population	Pakistani Population	Bangladeshi Population	Chinese Population	Other Asian Population	Total Asian Population	Neighborhood
8	Barnet	27920	5344	2215	8259	22180	65918	Barnet
3	Tower Hamlets	6787	2442	81377	8109	5786	104501	Tower Hamlets
16	Greenwich	7836	2594	1645	5061	12758	29894	Greenwich
5	Ealing	48240	14711	1786	4132	31570	100439	Ealing
14	Wandsworth	8642	9718	1493	3715	9770	33338	Wandsworth
2	Brent	58017	14381	1749	3250	28589	105986	Brent
7	Hillingdon	36795	9200	2639	2889	17730	69253	Hillingdon
19	Kingston Upon Thames	6325	3009	892	2883	13043	26152	Kingston Upon Thames
4	Harrow	63051	7797	1378	2629	26953	101808	Harrow
11	Merton	8106	7337	2216	2618	15866	36143	Merton
6	Hounslow	48161	13676	2189	2405	20826	87257	Hounslow

Cluster 3

	London Borough	Indian Population	Pakistani Population	Bangladeshi Population	Chinese Population	Other Asian Population	Total Asian Population	Neighborhood
18	Southwark	5819	1623	3912	8074	7764	27192	Southwark
12	Camden	6083	1489	12503	6493	8878	35446	Camden
15	Westminster	7213	2328	6299	5917	10105	31862	Westminster
13	Enfield	11648	2594	5599	2588	12464	34893	Enfield
10	Waltham Forest	9134	26347	4632	2579	11697	54389	Waltham Forest
17	Barking and Dagenham	7436	8007	7701	1315	5135	29594	Barking and Dagenham

Cluster 4

	London Borough	Indian Population	Pakistani Population	Bangladeshi Population	Chinese Population	Other Asian Population	Total Asian Population	Neighborhood
9	Croydon	24660	10865	2570	3925	17607	59627	Croydon

Cluster 5

	London Borough	Indian Population	Pakistani Population	Bangladeshi Population	Chinese Population	Other Asian Population	Total Asian Population	Neighborhood
0	Newham	42484	30307	37262	3930	19912	133895	Newham

Best cluster according with Chinese population

- As shown in the previous slides there are more Chinese population in the Cluster 2. The neighborhoods which lead the rank is Barnet and Tower Hamlets.

Group by neighborhood and check the mean value of each venue.

	Neighborhood	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Garage	Auto Workshop	Bagel Shop	Bakery	Chinese Restaurant	Chocolate Shop	Clothing Store	Cocktail Bar	Coffee Shop	Concert Hall	Construction & Landscaping
0	Barking and Dagenham	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.50
1	Barnet	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
2	Brent	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
3	Camden	0.037037	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.074074	0.000000	0.000000	0.000000	0.000000	0.074074	0.000000	0.00
4	Croydon	0.000000	0.000000	0.000000	0.035088	0.000000	0.000000	0.0	0.00	0.017544	0.000000	0.000000	0.070175	0.000000	0.175439	0.000000	0.00
5	Ealing	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.05	0.050000	0.000000	0.000000	0.000000	0.000000	0.100000	0.000000	0.00
6	Enfield	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.2	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.200000	0.000000	0.00
7	Greenwich	0.000000	0.000000	0.000000	0.045455	0.000000	0.000000	0.0	0.00	0.000000	0.090909	0.000000	0.000000	0.000000	0.045455	0.000000	0.00
8	Harrow	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000000	0.142857	0.000000	0.178571	0.000000	0.00
9	Hillingdon	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.250000	0.000000	0.000000	0.000000	0.000000	0.000000	0.25
10	Hounslow	0.000000	0.000000	0.000000	0.000000	0.000000	0.166667	0.0	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
11	Kingston Upon Thames	0.000000	0.012821	0.012821	0.000000	0.000000	0.000000	0.0	0.00	0.038462	0.012821	0.012821	0.051282	0.000000	0.089744	0.000000	0.00
12	Merton	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.071429	0.000000	0.000000	0.000000	0.000000	0.071429	0.000000	0.00
13	Newham	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
14	Redbridge	0.000000	0.000000	0.000000	0.000000	0.058824	0.000000	0.0	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.058824	0.000000	0.00
15	Southwark	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00
16	Tower Hamlets	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.100000	0.000000	0.00
17	Waltham Forest	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.133333	0.066667	0.00
18	Wandsworth	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.142857	0.000000	0.00
19	Westminster	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.000000	0.000000	0.017544	0.017544	0.122807	0.000000	0.00

- As shown in the “Group by neighborhood and check the mean value of each venue” there is no Chinese restaurant in the Neighborhoods with more number of Chinese population.

Conclusion

- As we can see, analyzing this dataset, the highest amount of Chinese population is located at Barnet and Tower Hamlets represented in cluster 2. Interestingly, we don't see any Chinese restaurants in the top ten most common venues. Therefore, we have a great business opportunity, in view of the large number of Chinese population living in this region