Optimal Location for A Restaurant

Jason Dunlop

Scenario - Analyzing a large city to identifying the best neighborhood to open a restaurant by:

- **Data Preparation:** Using open source data available online (Wikipedia) to pull city (London) population by ethnicity.

- **Data Understanding**: Using Foursquare data to identify the latitude and longitude coordinates of the neighborhoods that house the most popular restaurant doors

- **Data Understanding:** Clustering the neighborhoods by similarities in their venues

- **Deployment:** Analyze the results and make a conclusion based on the analysis.

Data Preparation

Data used to discover the population by Ethnicity to see if a specific Indian Cuisine could house the demand!

Out[2]:

|	London Borough	Indian Population	Pakistani Population	Bangladeshi Population	Chinese Population	Other Asian Population	Total Asian Population
0	Newham	42484	30307	37262	3930	19912	133895
1	Redbridge	45660	31051	16011	3000	20781	116503
2	Brent	58017	14381	1749	3250	28589	105986
3	Tower Hamlets	6787	2442	81377	8109	5786	104501
4	Harrow	63051	7797	1378	2629	26953	101808
5	Ealing	48240	14711	1786	4132	31570	100439
6	Hounslow	48161	13676	2189	2405	20826	87257
7	Hillingdon	36795	9200	2639	2889	17730	69253
8	Barnet	27920	5344	2215	8259	22180	65918
9	Croydon	24660	10865	2570	3925	17607	59627
10	Waltham Forest	9134	26347	4632	2579	11697	54389
11	Merton	8106	7337	2216	2618	15866	36143
12	Camden	6083	1489	12503	6493	8878	35446
13	Enfield	11648	2594	5599	2588	12464	34893
14	Wandsworth	8642	9718	1493	3715	9770	33338
15	Westminster	7213	2328	6299	5917	10105	31862
16	Greenwich	7836	2594	1645	5061	12758	29894
17	Barking and Dagenham	7436	8007	7701	1315	5135	29594
18	Southwark	5819	1623	3912	8074	7764	27192
19	Kingston Upon Thames	6325	3009	892	2883	13043	26152

Data Preparation

- Location data was appended to scrape the data in order access the Foursquare API

- The 2nd dataframe then organized by neighborhood to understand the relationships to each other in a large city

- Location coordinates added to the dataframe to help prepare for Mapping

Out[8]:

	Neighborhood	Latitude	Longitude
0	Newham	51.5255	0.0352
1	Redbridge	51.5901	0.0819
2	Brent	51.5673	-0.2711
3	Tower Hamlets	51.5203	-0.0293
4	Harrow	51.5806	-0.3420

Data Understanding

Using the location coordinates to show all the available
neighborhoods available to open restaurant location
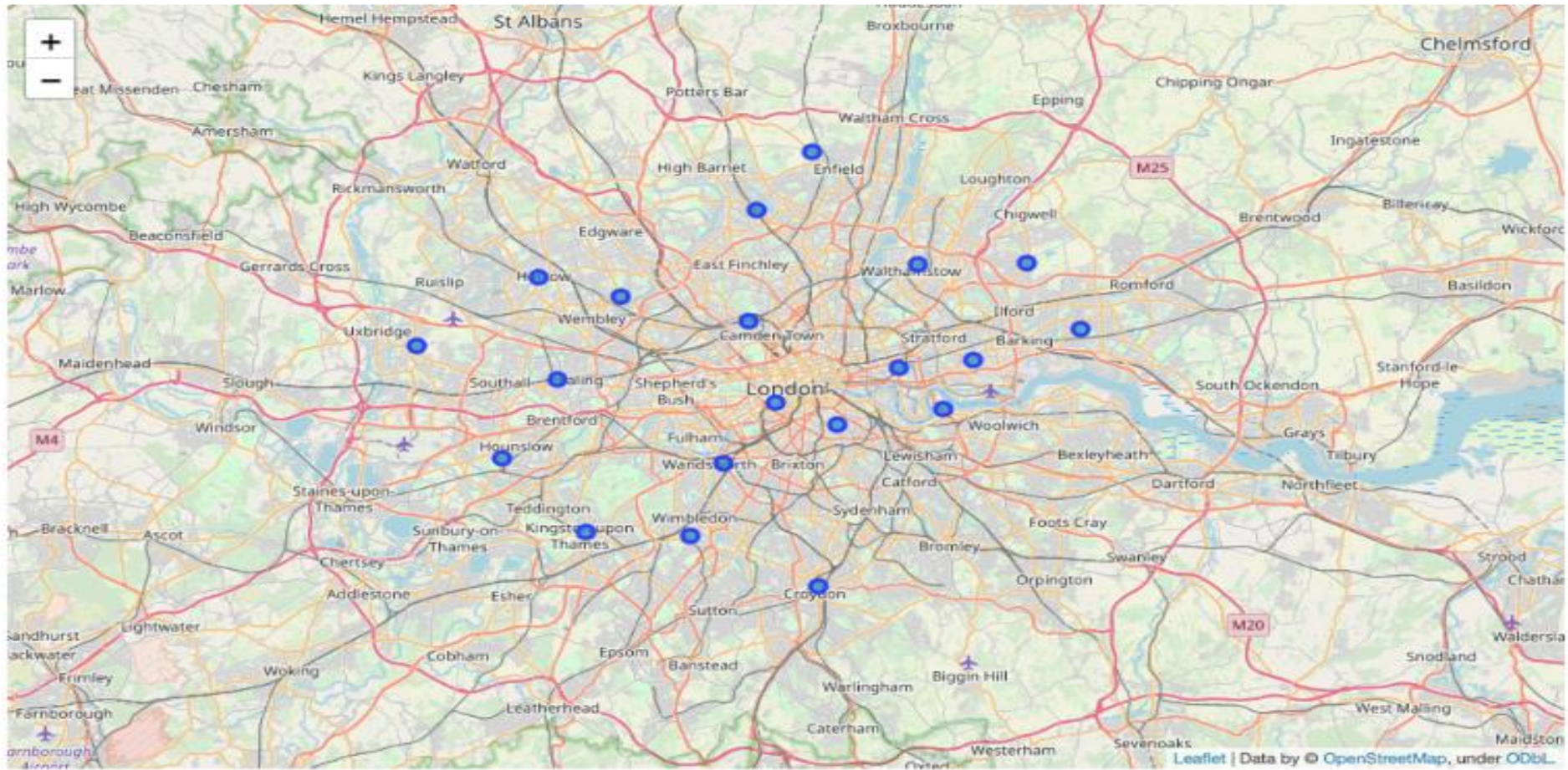
Data Understanding

Venue coordinates pulled within each neighborhood to show the ethnicity of the cuisine and their distance in relation to each other. Example below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Newham	51.5255	0.0352	Delicious Café	51.526417	0.030133	Café
1	Newham	51.5255	0.0352	Tesco Express	51.527187	0.035118	Grocery Store
2	Newham	51.5255	0.0352	Deep Blue Sea Fish & Chips	51.525097	0.039410	Fish & Chips Shop
3	Newham	51.5255	0.0352	West Ham Food Centre	51.527451	0.031644	Convenience Store
4	Newham	51.5255	0.0352	Ginny's Pie and Mash	51.525705	0.029532	Café

Venues frequency of occurrence by neighborhood to show the most popular cuisines. Example


```
----Barking and Dagenham----
                     venue  freq
0   Construction & Landscaping  0.33
1                        Lake  0.33
2                        Park  0.33
3          American Restaurant  0.00
4                   Nightclub  0.00
```

Data Understanding

Clusters were generated using K means

```
t_values(by = ['Indian Population'], ascending = False)
#Cluster 2
```

Out[43]:

	London Borough	Indian Population	Pakistani Population	Bangladeshi Population	Chinese Population	Other Asian Population	Total Asian Population	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Ve
13	Enfield	11648	2594	5599	2588	12464	34893	Enfield	Pub	Coffee Shop	Restaurant	Auto Works
14	Wandsworth	8642	9718	1493	3715	9770	33338	Wandsworth	Pub	Grocery Store	Café	Park
15	Westminster	7213	2328	6299	5917	10105	31862	Westminster	Hotel	Sandwich Place	Coffee Shop	Theat
12	Camden	6083	1489	12503	6493	8878	35446	Camden	Gastropub	Bakery	Pizza Place	Coffe Shop
18	Southwark	5819	1623	3912	8074	7764	27192	Southwark	Pub	Lake	Building	Café

In [44]:
```
london_merged.loc[london_merged['Cluster Labels'] == 2, london_merged.columns[[0] + list(range(1, london_merged.shape[1]))]].sor
t_values(by = ['Indian Population'], ascending = False)
#Cluster 3
```

Out[44]:

	London Borough	Indian Population	Pakistani Population	Bangladeshi Population	Chinese Population	Other Asian Population	Total Asian Population	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Mos Commo Venu
0	Newham	42484	30307	37262	3930	19912	133895	Newham	Grocery Store	Café	Convenience Store	Fish & Chips Shop

Data Understanding

Plotting clusters on a mapping for Data Visualization
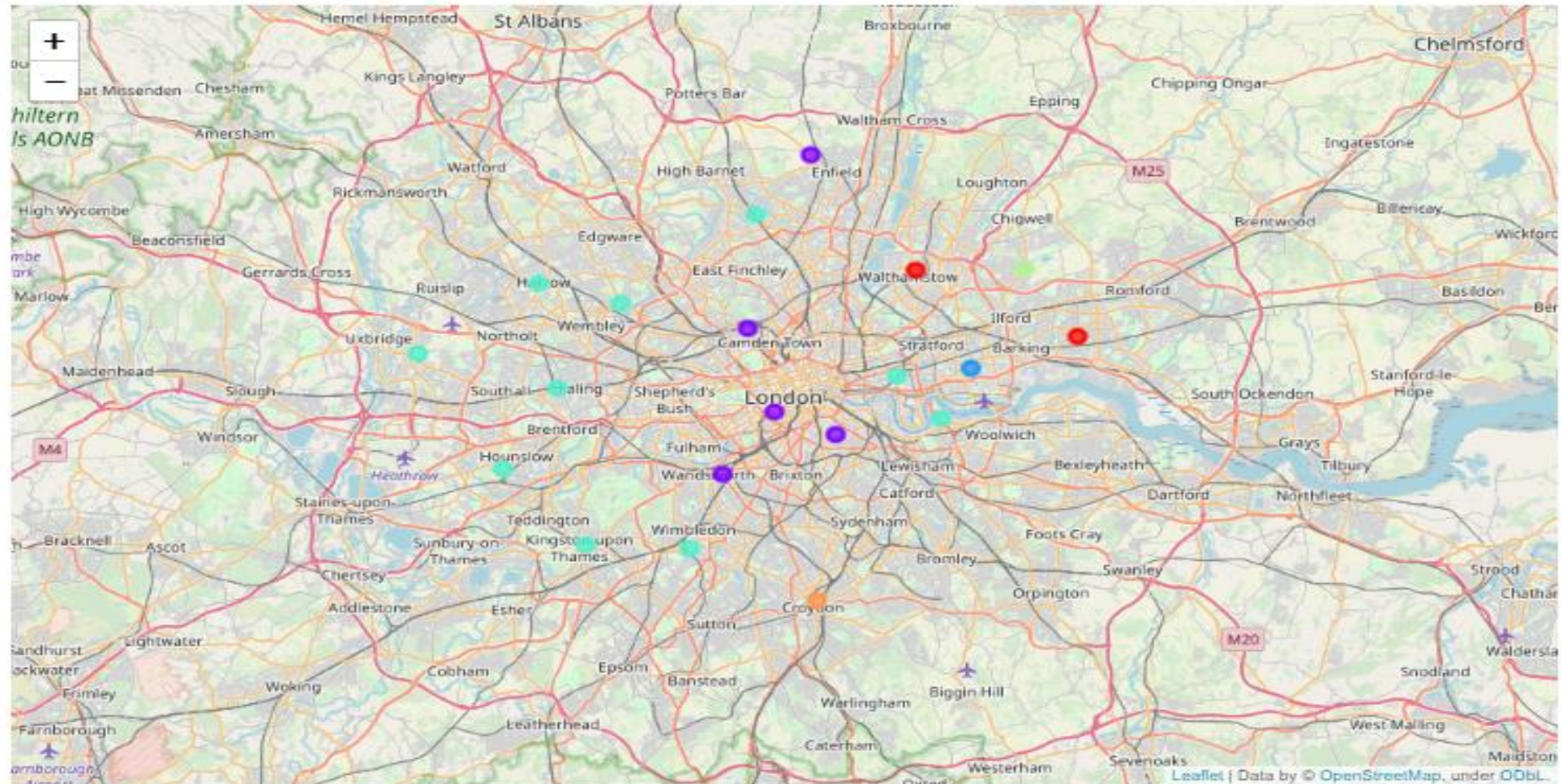
Conclusion

- As far as we can see with this data, the highest amount of Indian population is located at Harrow represented in cluster 1. If a deeper exam is performed into this cluster, it is noticeable that the living population in here ranks it the 5th most Asian inhabited borough. Apart of this fact, a strange closeness to Indian food can be found as the 8th most prominent venue in here is Middle Eastern restaurant which, while not being an Indian restaurant is the closest match to an Indian cuisine restaurant. By following this logic, if we would like to open a new Indian restaurant in the city or any kind of restaurant in fact, it would only be necessary to find a where are the restaurants similar the one we want to open, study the population in that area, and find similar clusters of population in the city that don't have yet or have very few restaurants like the one we would like to open. In this example, clusters 1 could make a good match for our target population. Looking at the venues in this cluster, it is possible to find two Indian restaurants, and a good bunch of Middle Eastern restaurants and coffee shops. So, in this cluster, it is possible to state that the existing restaurants matches the population's nationalities and tastes.

- In conclusion and taking into consideration the explanations given above as well as the data, it is highly possible that cluster 1 could be a good place to open our Indian restaurants. As explained above, the same logic could apply to open other kind of restaurants or business in any other area of the city. It is only necessary to examine the existing businesses in our target area, and study the population, then compare these 2 factors with the same ones in areas where there are existing business like the one we want to open, and then verify if the matching is correct.