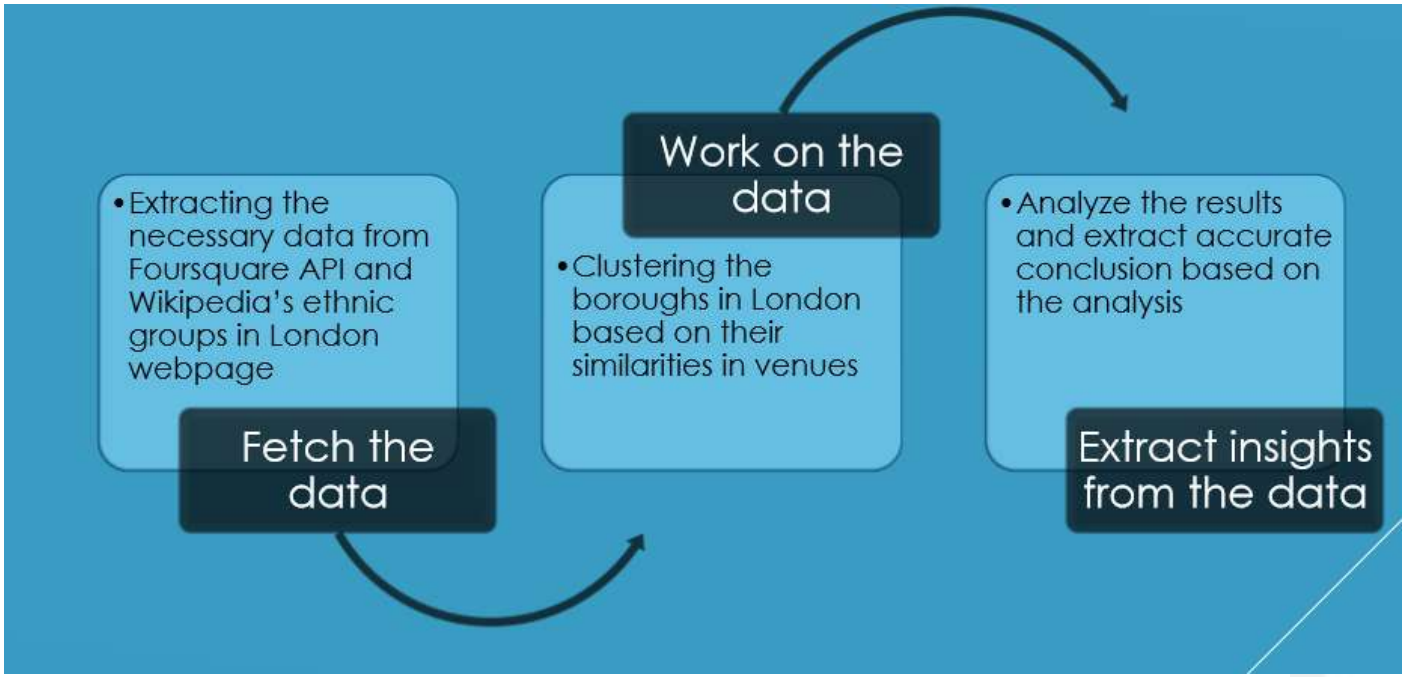
FINDING THE OPTIMAL LOCATION FOR A BUSINESS

Sergey Fotin

May, 2020

Applied Data Science Capstone

In this project we will determining location for starting a business in a city area by:



Applied Data Science Capstone

The data that was used contained the information about distribution of various Asian ethnicities with respect to their location in London:

Rank	London Borough	Indian Population	Pakistani Population	Bangladeshi Population	Chinese Population	Other Asian Population	Total Asian Population
1	Newham	42,484	30,307	37,262	3,930	19,912	133,895
2	Redbridge	45,660	31,051	16,011	3,000	20,781	116,503
3	Brent	58,017	14,381	1,749	3,250	28,589	105,986
4	Tower Hamlets	6,787	2,442	81,377	8,109	5,786	104,501
5	Harrow	63,051	7,797	1,378	2,629	26,953	101,808
6	Ealing	48,240	14,711	1,786	4,132	31,570	100,439
7	Hounslow	48,161	13,676	2,189	2,405	20,826	87,257
8	Hillingdon	36,795	9,200	2,639	2,889	17,730	69,253
9	Barnet	27,920	5,344	2,215	8,259	22,180	65,918
10	Croydon	24,660	10,865	2,570	3,925	17,607	59,627

Applied Data Science Capstone

Location data was appended to the scraped data in order to access Foursquare API.

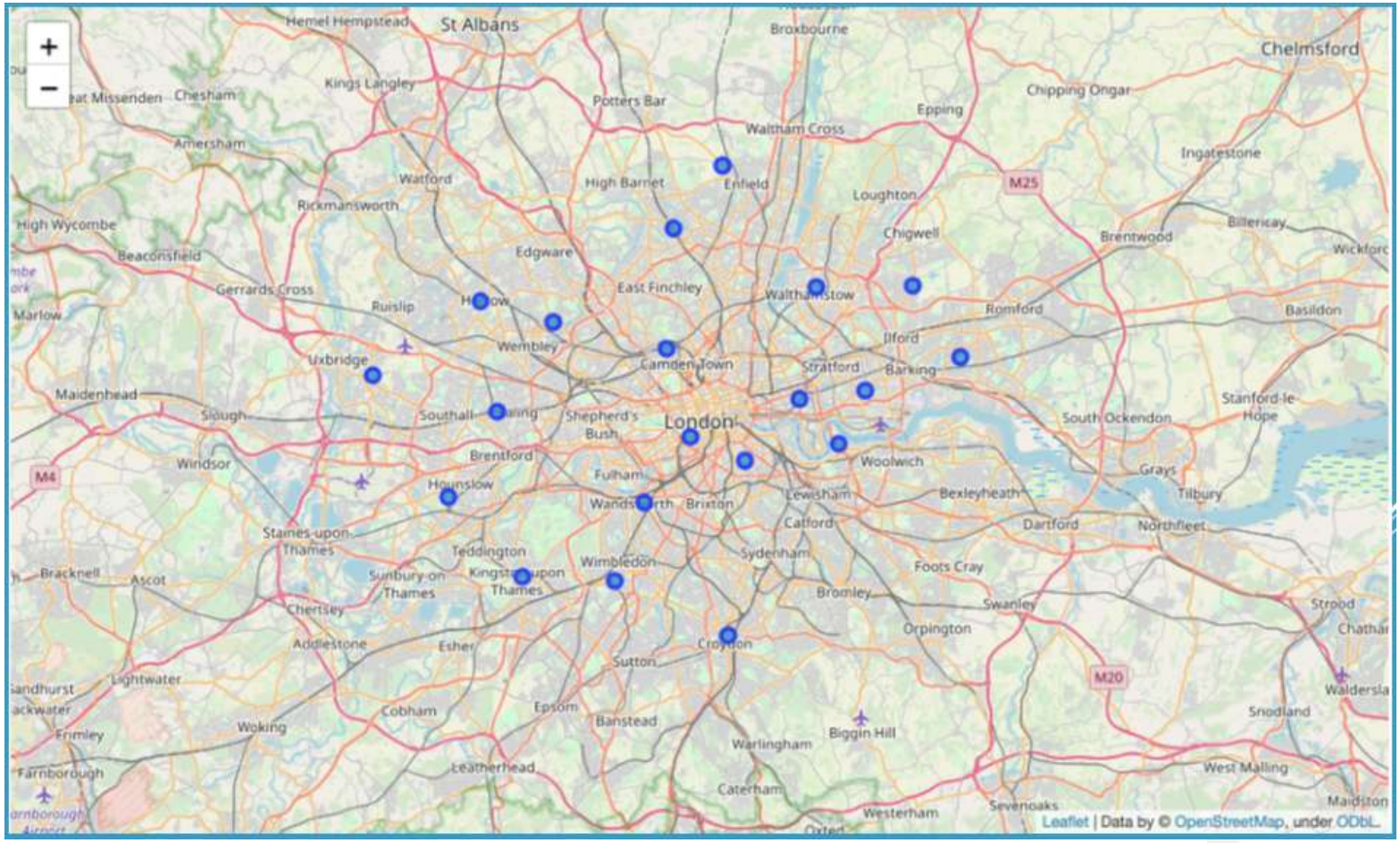
Second DataFrame was created with neighbourhood names.

Intel about latitude and longitude of each neighbourhood was added to this dataframe.

	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

Applied Data Science Capstone

A map was generated consisting of all the neighborhoods:



Applied Data Science Capstone

Nearby venues with respect to each neighborhood was extracted along with their frequencies of occurrences:

Venue's by neighborhood

	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	Andre Moves	51.524192	0.036145	Home Service
3	Newham	51.5255	0.0352	Deep Blue Sea Fish & Chips	51.525097	0.039410	Fish & Chips Shop
4	Newham	51.5255	0.0352	Ginny's Pie and Mash	51.525705	0.029532	Café

Venue's frequency of occurrence

----Barnet----

	venue	freq
0	Café	0.67
1	Bus Stop	0.33
2	American Restaurant	0.00
3	Recreation Center	0.00
4	Public Art	0.00

Applied Data Science Capstone

Clusters were generated using K-Means:

Cluster Three:

```
In [134]: london_merged.loc[london_merged['Cluster Labels'] == 2, london_merged.columns[[0] + list(range(1, london_merged.shape[1
```

Out[134]:

	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 Venue	5th Most Common Venue
15	Westminster	7213	2328	6299	5917	10105	31862	Westminster	Hotel	Coffee Shop	Sandwich Place	Sushi Restaurant	Theater

Cluster Four:

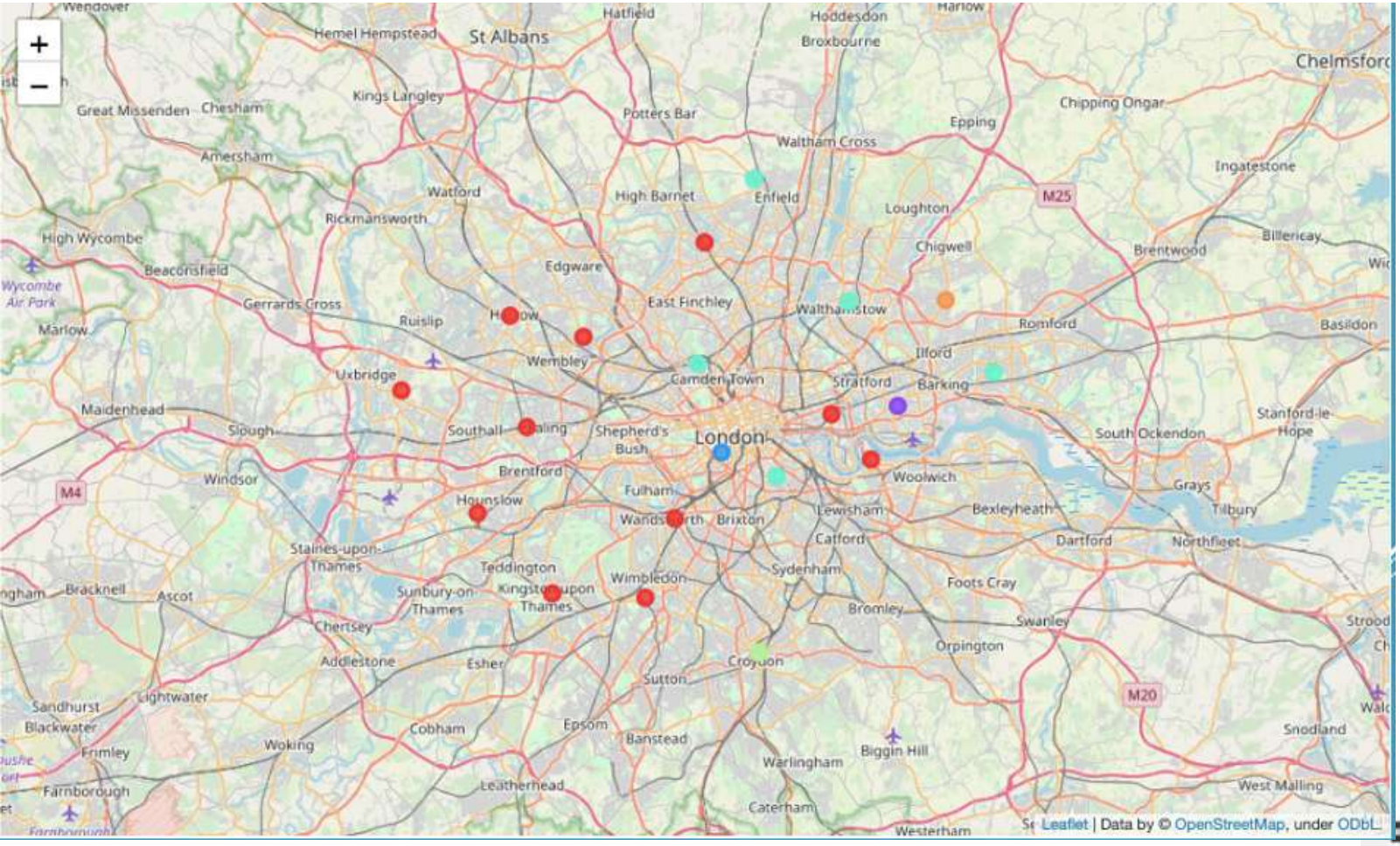
```
In [135]: london_merged.loc[london_merged['Cluster Labels'] == 3, london_merged.columns[[0] + list(range(1, london_merged.shape[1
```

Out[135]:

	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 Venue	5th Most Common Venue
13	Enfield	11648	2594	5599	2588	12464	34893	Enfield	Pub	Coffee Shop	Restaurant	Auto Workshop	Tennis Court
10	Waltham Forest	9134	26347	4632	2579	11697	54389	Waltham Forest	Grocery Store	Pub	Coffee Shop	Concert Hall	Vegetarian / Vegan Restaurant
17	Barking and Dagenham	7436	8007	7701	1315	5135	29594	Barking and Dagenham	Lake	Park	Women's Store	Food Court	Department Store
12	Camden	6083	1489	12503	6493	8878	35446	Camden	Gastropub	Bakery	Pizza Place	Coffee Shop	Café
18	Southwark	5819	1623	3912	8074	7764	27192	Southwark	Pub	Building	Café	Skate Park	Park

Applied Data Science Capstone

Clusters were plotted on the map for visualization:



Applied Data Science Capstone

CONCLUSION

•As far as we are able to 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.

Applied Data Science Capstone

CONCLUSION

•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.