1.1 Introduction:

London is a great city having people from diversed ethnicity. People have different preferences when it comes to select Neighborhood.

They prefer to live in neighborhood where they find some interest in places and some similarity in their culture or people. That is why every city have specific kind of area for each ethnic area. That way they interact with each other easily, they get their prferred food easily. This analysis is about find similarity in same so that result can be used for the problem discussed in description.

2.1 Description:

Lets take a scenario that we want to open an Indian restaurant in London but we are not sure about "WHERE". For this purpose, we will study some Neighborhood in London. We will try to cluster them, find similarity and then try to see the ranking of places where it will be better to open a restaurant.

This depends on a hypothesis that higher the number of Indian people larger is the opportunity. Also this opportunity decreases with existing Indian restaurant aas they also share customers for their food.

3.1 Data Collection:

Data Sources:

WE have used few Neighborhood available present at below wikipedia link. This gives us the Neighborhood and Indian people living in them.

https://en.wikipedia.org/wiki/Ethnic_groups_in_London

Few of the Neighborhood from this have been taken for study. Also we have used foursquare api to fetch the details of each neighborhood which we will study

3.2 Data Cleaning and feature engineering:

We have extractes the data from wikipedia which have only population and neighborhood. Later we fetched the latitude and longitude for them.

Also all the venues have been fetched from four square api. Later these were cleaned, grouped. New feature have been derived from this data.

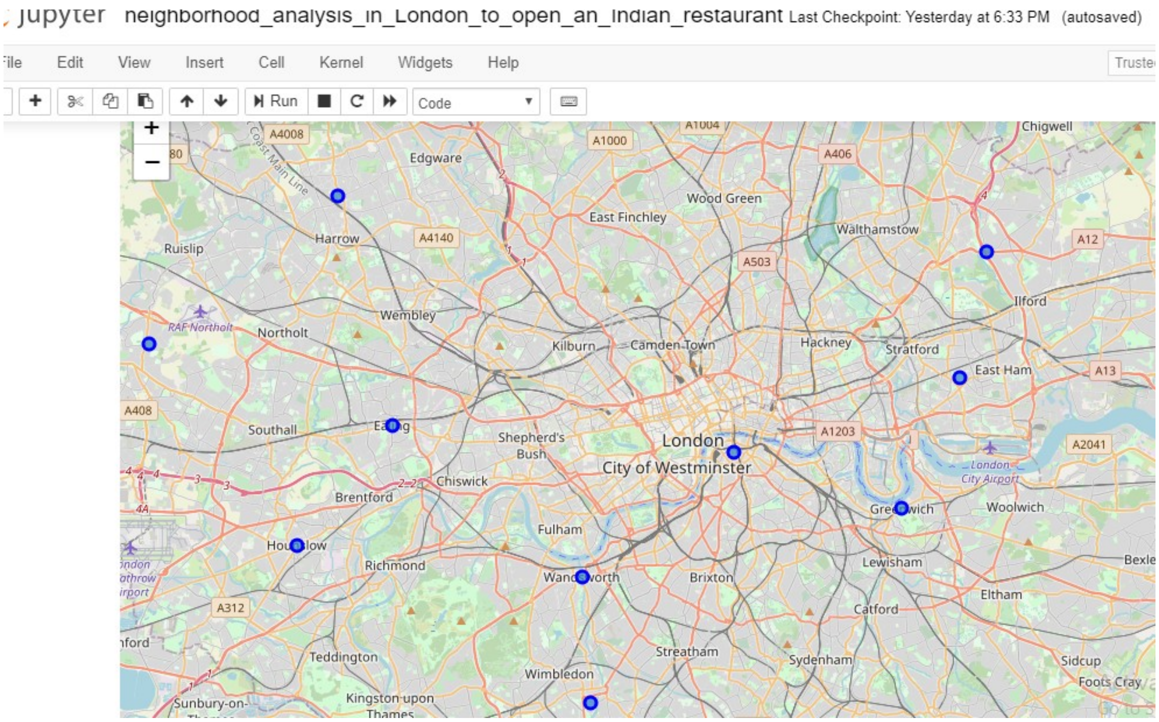
3.3 Feature Selection:

The features for neighborhood are the the number of food places, entertainment places, health centers like gym , cafes, hotels etc. For each venue, we fetched the different categories and later used all distinct values as feature. This required to study the data and change them in required format.

4.1 Exploratory Data Analysis

We used Folium to plot the data Neighborhood to see their location and confirm that we are covering enough area for deciding to open a restaurant.

For this we fetched the latitude and longitude of the Neighborhoods using Geolocator and later plot these locations on London map.



5.1 Predictive Modelling

For this we used knn model since we wanted to see the similarity in between neighborhoods based on their venues. We tried to divide the neighborhoods in 4 distinct category based on venues available at those locations.

4]:

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2 C
0	Newham	51.530000	0.029318	3	Café	
1	Redbridge	51.576320	0.045410	1	Hotel	
2	Brent	32.937346	-87.164718	0	Convenience Store	
3	Harrow	51.596769	-0.337275	1	Indian Restaurant	S
4	Ealing	51.512655	-0.305195	1	Coffee Shop	
5	Hounslow	51.468613	-0.361347	1	Fast Food Restaurant	
6	Hillingdon	51.542519	-0.448335	3	Park	F Re
7	Barnet	51.653090	-0.200226	1	Coffee Shop	Re
8	Croydon	51.371305	-0.101957	1	Pub	
9	Merton	51.410803	-0.188099	1	Tram Station	

6.1 Conclusion:

We were able to group the neighborhoods in different clusters. Also we produced a list which shows the ranking of Neighborhood in terms of recommendation to open an Indian restaurant in Neighborhoods.

```
df_rank
```

	Neighborhood	Indian Restaurant	normalized_population	Cluster Labels	ranking
3	Croydon	0.041667	1.000000	1	0.791667
2	Camden	0.000000	0.920160	0	0.736128
4	Ealing	0.010638	0.765095	1	0.609948
5	Enfield	0.016129	0.763842	1	0.607848
1	Brent	0.000000	0.724176	1	0.579341
0	Barnet	0.030303	0.673804	3	0.532982
6	Greenwich	0.000000	0.583575	3	0.466860
7	Harrow	0.125000	0.442816	1	0.329253
8	Hillingdon	0.000000	0.391112	1	0.312890
11	Newham	0.000000	0.184739	1	0.147791
12	Redbridge	0.000000	0.137064	1	0.109651

Thank you.