

Battle of Neighborhood

London, UK

1. Introduction: Business Problem

People often explore the neighborhood before they want to move to any specific location to understand various aspect of the locations - rentals, near-by shops, clinics / hospitals, schools, safety, distance to the office etc.

In this project, I have chosen to explore London neighborhood, where most of the Asian / Indian community reside and the surrounding restaurants around that area using clustering & segmentation techniques learned in this course. This project will also help in comparing 2 different neighborhoods to choose the best suited location based on the top 10 common venues surrounding it.

Objective:

- Extract top trending venues of London using Foursquare API
- Forming neighborhood clusters based on venue categories using unsupervised k-mean clustering algorithm
- Understanding the similarities and differences between two neighborhoods to retrieve more insights and to conclude which neighborhood is best suited for an individual's need.

2. Datasets and APIs:

DataSet1 - List of areas in London

I have choose to extract the list of areas in London from this wikipage:
https://en.wikipedia.org/wiki/List_of_areas_of_London

BeautifulSoap and other panda libraries were used to cleans this data for the required Neighborhood (with postal code) information of London.

DataSet2 - Demography of London

I have also used the demographic data of London from this wikipage:
https://en.wikipedia.org/wiki/Demography_of_London to find the top Asian locality within London

Important Libraries & API used in the project - geocoder, Folium, kMeans, Foursquare

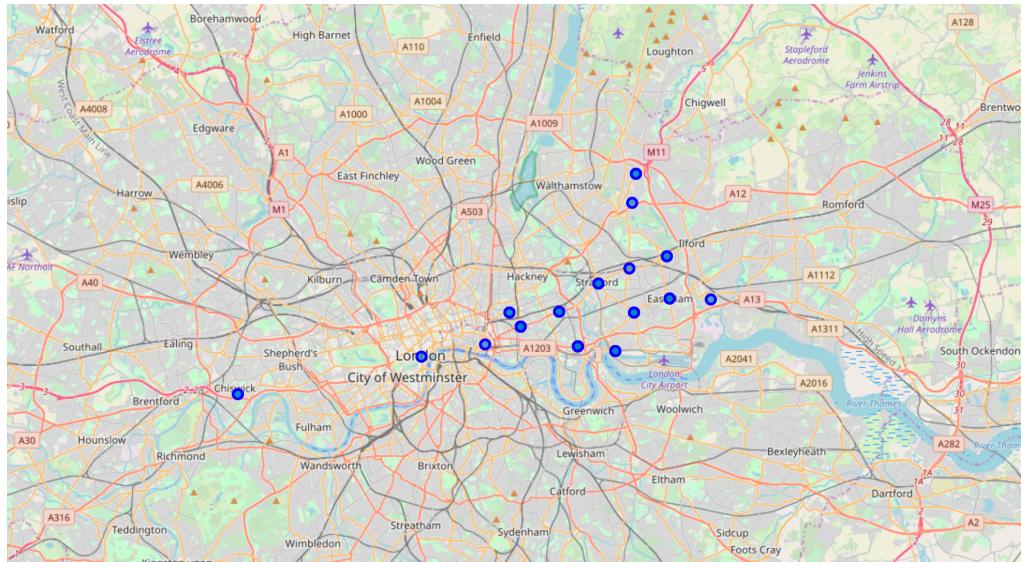
- Geocoder / Geopy library is used to fetch the required latitude and longitude information of each postal code of London, UK
- Folium is a great visualization library for plotting the map. This was used to visualize the neighborhoods cluster distribution of London city over an interactive leaflet map.
- Foursquare API was used to fetch the neighborhood of London and more specifically Newham borough to find the top common venues using the API.
- kMeans clustering was used to form the clusters of surrounding neighborhoods and compare the neighborhood information.

3. Methodology

After deciding on the dataset, went through the process of Web Scraping, Data Extraction and Data Wrangling using BeautifulSoup and other panda libraries to get the required London Neighborhood information in a clean dataset. Using geocoder, latitude and longitude information was successfully appended to the data set.

It is then merged with the second dataset which contains demographic information of London to choose the top 5 boroughs where Asian community resides.

Choropleth visualizing map of London is then plotted using folium map.



FourSquare API is used to retrieve the top trending venues of the neighborhoods and kMeans clustering techniques are used to cluster the neighborhood area.

4. Results

Using clustering techniques, top trending venues of each neighborhood were found within Newham. Example of “Beckton” and “Forest Gate” area can be seen below:

----Beckton----			----Forest Gate----		
	venue	freq		venue	freq
0	Hotel	0.07	0	Grocery Store	0.16
1	Coffee Shop	0.05	1	Fast Food Restaurant	0.11
2	Café	0.05	2	Bus Stop	0.11
3	Fast Food Restaurant	0.05	3	Pub	0.11
4	Clothing Store	0.05	4	Hotel	0.05
5	Grocery Store	0.05	5	Indian Restaurant	0.05
6	Bus Station	0.05	6	Fish & Chips Shop	0.05
7	Steakhouse	0.02	7	Market	0.05
8	Pier	0.02	8	Comfort Food Restaurant	0.05
9	Gym Pool	0.02	9	Chinese Restaurant	0.05

Newham Neighborhood and its most common venues were captured:

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0 Beckton	Hotel	Coffee Shop	Bus Station	Grocery Store	Café	Fast Food Restaurant	Clothing Store
1 Canning Town	Hotel	Athletics & Sports	Scenic Lookout	Grocery Store	Italian Restaurant	Middle Eastern Restaurant	Diner
2 Custom House	Hotel	Athletics & Sports	Scenic Lookout	Grocery Store	Italian Restaurant	Middle Eastern Restaurant	Diner
3 East Ham	Clothing Store	Warehouse Store	Sandwich Place	Grocery Store	Turkish Restaurant	Fast Food Restaurant	Electronics Store
4 Forest Gate	Grocery Store	Fast Food Restaurant	Bus Stop	Pub	Comfort Food Restaurant	Chinese Restaurant	Indian Restaurant

Comparison results of 2 Neighborhood to analyze and conclude, if it satisfies an individual's requirement where he/she decides to live:

Location	Forest Gate	Maryland
Borough	Newham	Newham
Post_town	LONDON	LONDON
PostalCode	E7	E15
Latitude	51.5467	51.54
Longitude	0.02558	0.00289
Cluster Labels	0	3
1st Most Common Venue	Grocery Store	Pub
2nd Most Common Venue	Fast Food Restaurant	Platform
3rd Most Common Venue	Bus Stop	Hotel
4th Most Common Venue	Pub	Sandwich Place
5th Most Common Venue	Comfort Food Restaurant	Supermarket
6th Most Common Venue	Chinese Restaurant	Coffee Shop
7th Most Common Venue	Indian Restaurant	Bookstore
8th Most Common Venue	Market	Café
9th Most Common Venue	Fish & Chips Shop	General Entertainment
10th Most Common Venue	Moving Target	Indoor Play Area

5. Discussion

This will be of great benefit to many individuals / families who move to different localities / countries for various reasons – be it job, business or migration. This could be a recommender system to investigate / explore / analyze the choice of neighborhood where people want to live.

6. Conclusion

This Analysis concludes that compared to Maryland, Forest Gate satisfies all the requirements:

- Both satisfies the Asian community since its part of Newham borough
- Top 10 common venues shows Forest Gate has got a good neighborhood with the requirement such as – nearby Market / Grocery shops, Bus stop / Metro, and Indian / Chinese Restaurants in the locality.