Best Location in London for Chinese Restaurant

1. Introduction

1.1 Background

In this capstone project of applied data science course, we shall cluster and compare the neighborhood in London to help us find the best location to open a restaurant. Based on the similarity of popular venues around in the area.

1.2Problem description

One of my friends, Jimmy, is running a successful Chinese restaurant in Soho,London. He is looking to expand his business by opening another restaurant. We will analyse the constitute of venues in each neighborhood in London to find the most similar neighborhood to Soho. I believe this will increase the chance of success.

1.3.Target Audience

Originally this project was only designed for Jimmy to find the best location to open a new restaurant. But it can be helpful to other businesses to locate their business.

2.Data

2.1 We can find the list of London in London from Wikipedia
https://en.wikipedia.org/wiki/List_of_areas_of_London

2.2 Extract all the Location name from the dataset and find the latitude and longitude

2.3 Use Foursquare API to gather the venues' information such as name, category, latitude,longitude ect. in each Location.

3. Methodology

3.1 Data Retrieval,Cleaning and Feature Engineering

The London location information is from Wikipedia, in csv format, contains 5 columns

1. Location
2. London borough
3. Post town
4. Postcode district
5. Dial code
6. OS grip ref

	Location	London borough	Post town	Postcode district	Dial code	OS grid ref
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	20	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	20	TQ205805
2	Addington	Croydon[8]	CROYDON	CR0	20	TQ375645
3	Addiscombe	Croydon[8]	CROYDON	CR0	20	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	20	TQ478728

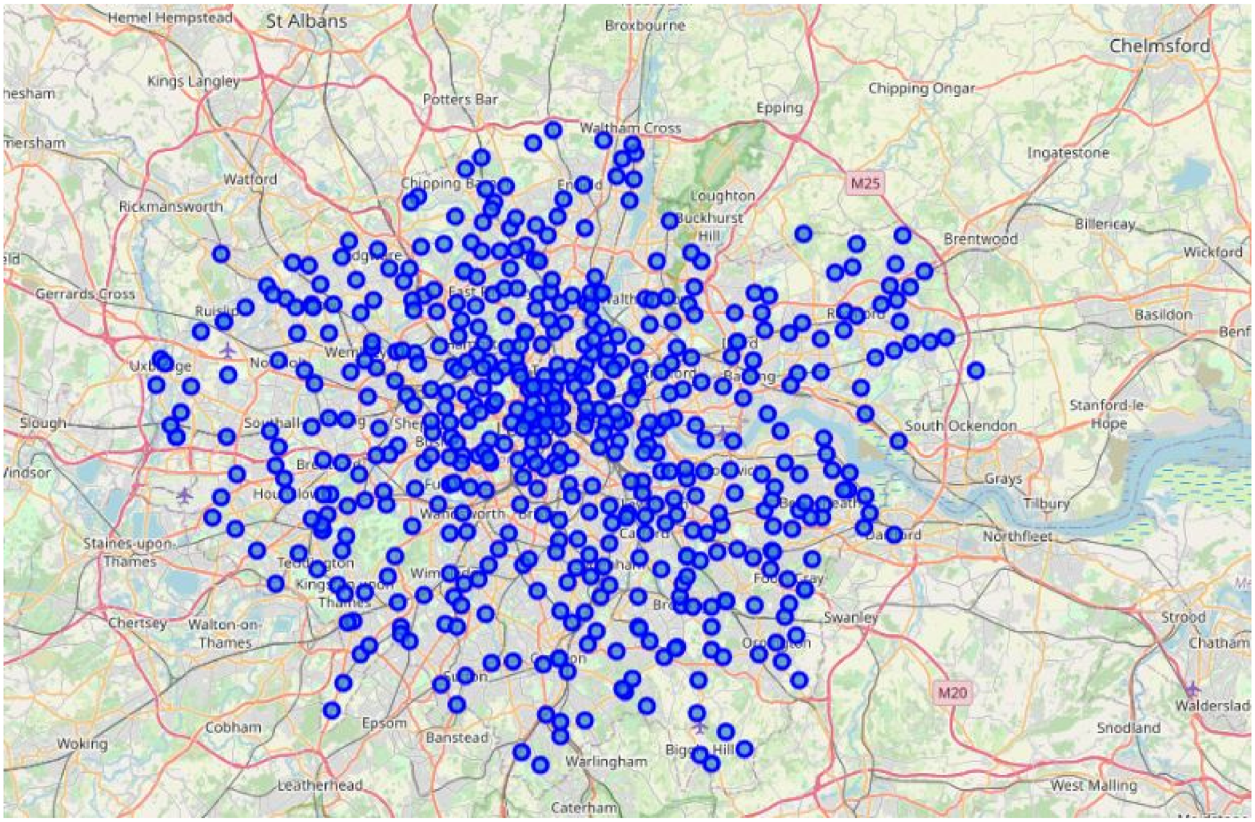
Location and Post town columns are extracted for further analysis.there are 533 locations in London

	Location	Post town	Latitude	Longitude
0	Abbey Wood	LONDON	51.487621	0.114050
1	Acton	LONDON	51.508140	-0.273261
2	Addington	CROYDON	51.358637	-0.031635
3	Addiscombe	CROYDON	51.379692	-0.074282
4	Albany Park	BEXLEY, SIDCUP	51.423709	0.099809

Then latitudes and longitudes of each location was fetched using **Geocoder** from Geopy package. It was observed that Nominatim geocoder was not able to give location data for all the Location in the dataset. After trail,40 locations need be checked and entered manually from Google Map.

3.2 Exploratory Data Analysis\

Plot London locations on map using Nominatim and Folium.



Explore different locations by fetching 100 top venues within a radius of 500 m using **Foursquare** api.

Defined a custom function get_nearby_venues to get Venue, Venue Latitude, Venue Longitude, Venue Category for each location in London Dataset. Applied the function to all rows of London dataset.

	Location	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Abbey Wood	Grocery Store	Playground	Campground	Zoo Exhibit	Farm	Farmers Market	Fast Food Restaurant	Filipino Restaurant	Film Studio	Fish & Chips Shop
1	Acton	Pub	Gym / Fitness Center	Chinese Restaurant	Coffee Shop	Creperie	Bus Stop	Wine Shop	Fast Food Restaurant	Brewery	Grocery Store
2	Addington	Tram Station	Bus Station	English Restaurant	Park	Zoo Exhibit	Food	Farmers Market	Fast Food Restaurant	Filipino Restaurant	Film Studio
3	Addiscombe	Grocery Store	Bakery	Park	Pub	Fast Food Restaurant	Chinese Restaurant	Cosmetics Shop	Diner	Café	Fish & Chips Shop
4	Albany Park	Pharmacy	Fast Food Restaurant	Coffee Shop	Grocery Store	Café	Hotel	Pizza Place	Bakery	Pub	Italian Restaurant

Plot the number of venues for each neighbourhood. We can that there is huge variation in number of venues for each location in London

Converting venue category in numeric variables using One Hot Encoding(pandas dummy variables) and then grouping category-wise total venues for each location.

3.2 Clustering London Location

First of all, we cluster the locations within the city of London using kmeans. Initially, kmeans was run with parameters values:

1. n_clusters=kclusters (10)

2. random_state=0,

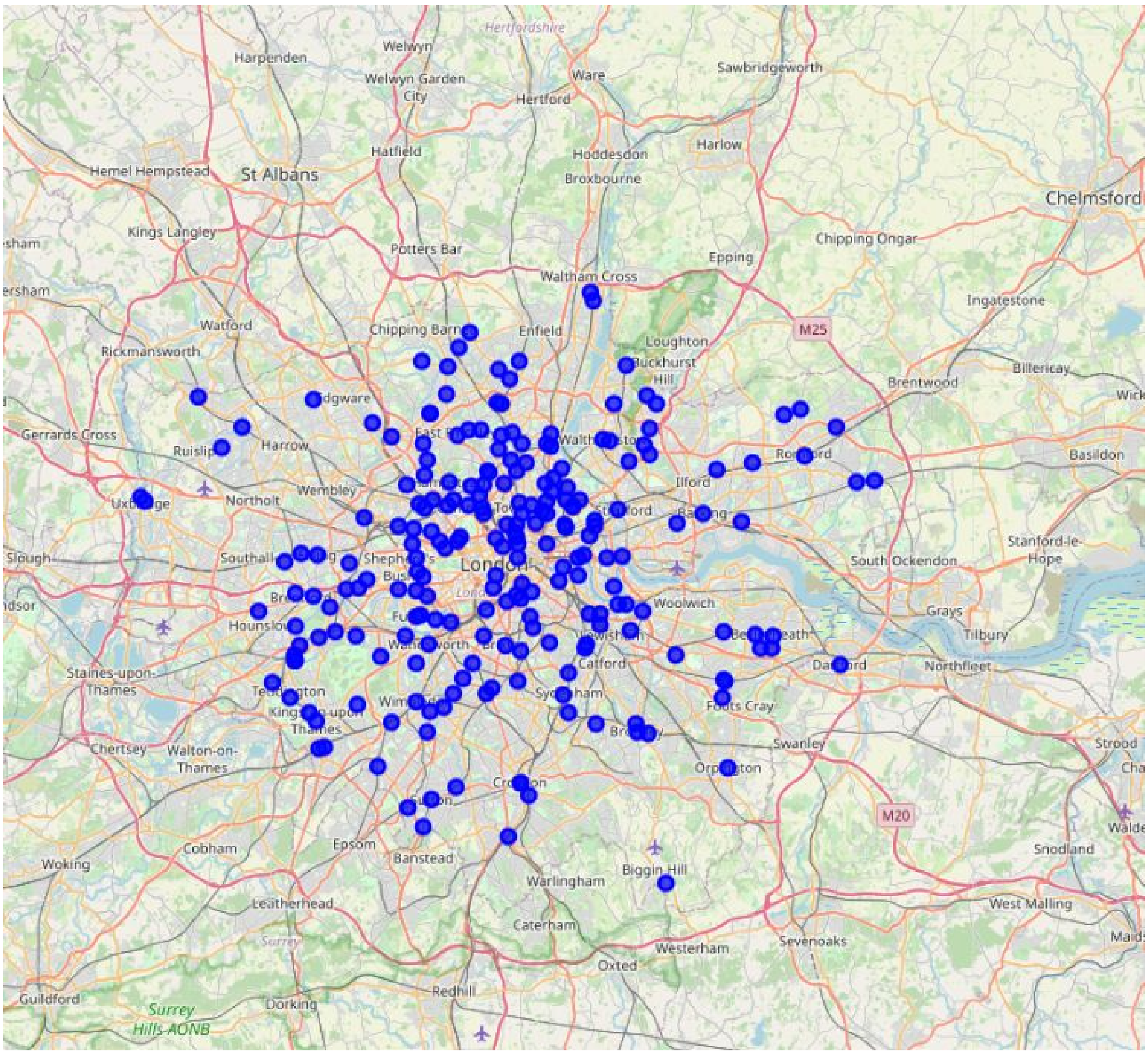
3. n_init = N_INIT(20)

4. algorithm='auto'

After clustering, the cluster labels were merged in the London location dataset . **the locations in the same cluster with soho is selected.**

3.3 Assort location with frequency of Chinese restaurant

We sort the selected location dataframe sort in ascending order by the frequency of Chinese restaurants .



4. Conclusion

In this project, we tried to find the location in London which is similar to soho, the make the best adaption to suit Jimmy's new restaurant .We used Foursquare API to analyse the types of businesses in each location. Use the Kmean method to cluster all the locations. After machine learning we get a list of locations that fall into the same cluster with soho, those locations has similirest business environment with soho. Then we found the location with most Chinese restaurants frequency. The location with more Chinese restaurant occupied means the population in this location likes Chinese better.

5.Further Study

Choosing a location for a business is stressful work and is the foundation of a successful business. In this project we focused on the similarity of different locations. We may need to consider more points such as the demographic of the population. The ethnic group of population and so on.

Thank you