

– The Battle of the Neighborhoods –

Exploring London and Identifying the Best Neighborhood to Open a New Chinese Restaurant

1. Introduction

1.1. Background

London is one of the biggest cities in the world with a population of over 8 million people, as per a Census performed in 2011. Subsequently, the city has attracted individuals from all over the world. It has become one of the most ethnically diverse cities in the world. London is melting pot of various people, ethnicities, cultures and backgrounds. Asians are one of its largest minority ethnic group. As per the 2011 Census, Asians make up 18.4% of the total population in London. This includes many Chinese, both immigrants and those that were born and raised in the UK. As such, it is no surprise that Chinese cuisine has been increasing in popularity. London even has its very own Chinatown located in Westminster.

1.2. Business Problem

Our business problem here is to capitalize on this increasing demand and interest in Chinese cuisine and open a Chinese restaurant. However, the first thing to think about when opening a new restaurant is location. The purpose of this project is to determine which neighborhood would be the most ideal to open a new Chinese restaurant. To do so, we will be analyzing the demographic data of boroughs in London and nearby venues as well as performing clustering on neighborhoods.

1.3. Who would be interested?

Any person/company looking to open a new Chinese restaurant in London.

2. Data acquisition and cleaning

2.1. Data Sources

We will be using 3 main data sources:

- Firstly, we want to figure out which boroughs have the highest Chinese population since one of the assumptions we will be making in this project is that the demand for Chinese food is predominantly from the Chinese community. Thus, we need demographic data in London. We can obtain this information from the [Wikipedia page – Ethnic Groups in London](#).
- Next, we will be choosing the top five boroughs with the highest Chinese population and obtain a list of neighborhoods and respective postal codes for those boroughs. The postal codes will then be used to get the coordinates (latitude and longitude) of the neighborhood which we will need later on. We can obtain this information from the [Wikipedia page – List of Areas of London](#).
- Lastly, we need to use the FourSquare API to explore each neighborhood and analyze what venues are in each neighborhood, how many Chinese restaurants there are, and which venues are the most common in each neighborhood. This will guide our final

recommendation on the neighborhood that is the most suitable to open a new Chinese restaurant. We can obtain this information from the [Foursquare Developer API](#).

2.2. Data Cleaning

We start off by webscraping the Wikipedia page, Ethnic Groups in London. We identify the table relevant to our project which is titled Asian Population of London. The html table is converted into a pandas dataframe. We only extract the two relevant columns, London Borough and Chinese Population. We then sort the dataframe in descending order by Chinese Population to get the top 5 boroughs with the highest Chinese population.

Next, we need to get the neighborhoods and postal codes for the 5 boroughs. We webscrape the Wikipedia page, List of Areas of London. From the html table, we extract the relevant columns – Location, London borough, Post town, Postcode district – into a new pandas dataframe. We note that in some cases the neighbourhood is affixed to more than one postcode. Since we only need one postcode to extract the coordinates of the neighbourhood, we select the first postcode listed and drop the rest. After that, we filter the column post town for only values that contain 'LONDON'. Then, we filter the resulting dataframe for the 5 London boroughs that we identified earlier.

3. Methodology

The main dataframe we will be using contains the following data: Neighborhood, Borough, PostCode, Latitude and Longitude.

	Neighborhood	Borough	PostCode	Latitude	Longitude
0	Aldwych	Westminster	WC2	51.5142	-0.123382
1	Arkley	Barnet	EN5	51.6562	-0.194317
2	Arkley	Barnet	NW7	51.6143	-0.227300
3	Bankside	Southwark	SE1	51.4963	-0.093038
4	Barnet Gate	Barnet	NW7	51.6143	-0.227300

3.1. Selecting the boroughs to analyze

There are over 300 neighborhoods in London. To reduce the amount of data processing, we will focus our analysis on **the top 5 boroughs with the highest Chinese population**. We assume that the Chinese population make up the majority of the market for Chinese cuisine. In reality, other ethnicities also do frequent Chinese restaurants. However, since market data for Chinese cuisine is not made freely available on the internet, for the purpose of this project, we will make this assumption.

As stated in 2.2, we determine the top 5 boroughs by filtering the Chinese Population column in descending order. We end up with **54** neighborhoods.

	London Borough	Chinese Population
0	Barnet	8259
1	Tower Hamlets	8109
2	Southwark	8074
3	Camden	6493
4	Westminster	5917

3.2. Exploring a Sample Neighborhood

We shall do an initial exploration of a sample neighborhood, Arkley, Barnet, to determine the workability of the FourSquare API data.

We perform a GET request to the FourSquare API, using a limit of 100 and a radius of 2000, and examine the initial results. There are 42 venues in Arkley, Barnet.

We drill into the data by generating a list of all 42 venues and their respective categories, i.e., pub, Italian restaurant, etc. We end up with the below dataframe.

	name	categories	address	lat	lng	labeledLatLngs	distance	postalCode	cc	city	state	country	formattedAddress
0	Ye Old Mitre Inne	Pub	58 High Street	51.652940	-0.199507	[[{"label": "display", "lat": 51.65293985597287...	510	EN5 5SJ	GB	Hertfordshire	Hertfordshire	United Kingdom	[58 High Street, Hertfordshire, EN5 5SJ, Unite...
1	The Black Horse	Pub	Wood St	51.653075	-0.206719	[[{"label": "display", "lat": 51.65307467634626...	924	EN5 4BW	GB	London	Greater London	United Kingdom	[Wood St, London, Greater London, EN5 4BW, Uni...
2	Everyman Cinema	Movie Theater	Great North Rd	51.646793	-0.187675	[[{"label": "display", "lat": 51.64679349064748...	1143	EN5 1AB	GB	Barnet	Greater London	United Kingdom	[Great North Rd, Barnet, Greater London, EN5 1...
3	Joie de Vie	Bakery	NaN	51.653659	-0.201288	[[{"label": "display", "lat": 51.653659, "lng": ...	558	NaN	GB	NaN	NaN	United Kingdom	[United Kingdom]
4	Caff� Nero	Coffee Shop	128 High St	51.654861	-0.201743	[[{"label": "display", "lat": 51.65486135090324...	534	EN5 5XQ	GB	Barnet	Greater London	United Kingdom	[128 High St, Barnet, Greater London, EN5 5XQ,...

Now we want to check how many Chinese restaurant, if any, there are in this neighborhood. To do this, we create a loop iterating through the above dataframe. If the categories column is 'Chinese Restaurant', we will print the row. Interestingly, after running the loop, the output is empty. This means that Arkley, Barnet does not have any Chinese restaurants in the area although it is in the borough with the highest Chinese population.

3.3. Exploring all the Neighborhoods in the 5 Boroughs

Now let us explore all the neighborhoods within the 5 boroughs, Barnet, Tower Hamlets, Southwark, Camden and Westminster, and determine which neighborhoods have the most Chinese restaurants. By analyzing which neighborhoods have the most Chinese restaurants, we can indirectly gauge where the demand for Chinese cuisine is.

Firstly, as above, we extract all the selected neighborhoods' venues and their respective categories. We then run a one hot encoding on the results, group the results by neighborhood and calculate the frequency of each venue category. We end up with the following table.

Neighborhood	Yoga Studio	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Art & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	Austrian Restaurant	BBQ Joint
0	Aldwych	0.0	0.00	0.000000	0.0	0.000000	0.00	0.0	0.01	0.0	0.0	0.0	0.0
1	Arkley	0.0	0.00	0.000000	0.0	0.000000	0.00	0.0	0.00	0.0	0.0	0.0	0.0
2	Bankside	0.0	0.00	0.032258	0.0	0.032258	0.00	0.0	0.00	0.0	0.0	0.0	0.0
3	Barnet Gate	0.0	0.00	0.000000	0.0	0.000000	0.00	0.0	0.00	0.0	0.0	0.0	0.0
4	Bayswater	0.0	0.01	0.010000	0.0	0.000000	0.01	0.0	0.00	0.0	0.0	0.01	0.0

We are only interested in the frequency of Chinese restaurants, so we filter out those neighborhoods where the frequency of Chinese restaurants is 0. Then, we drop all columns other than 'Neighborhood' and 'Chinese Restaurant'. Finally, we sort the result in descending order. Our new dataframe is as follows.

	Neighborhood	Chinese Restaurant
11	Brent Cross	0.200000
17	Colindale	0.142857
10	Bow	0.066667
35	Nunhead	0.062500
16	Church End	0.034483
39	Rotherhithe	0.034483
23	Finchley	0.030303
4	Bayswater	0.030000
31	Little Venice	0.030000
9	Bloomsbury	0.020833
13	Camberwell	0.018519

We have narrow down our possible neighborhood options to **11 neighborhoods**. But we want more insights into these neighborhoods before we make our conclusions and recommendations. One thing we can do is check what are the **10 most common venues in each neighborhood** to determine if Chinese restaurants is one of them and also where it places in the 10 most common venues. Using the one hot encoding dataframe, we create a function that will return the most common venues and set the value as 10. We generate a new dataframe with the results.

3.4. Clustering the Neighborhoods

We want to cluster the neighborhoods to determine if there is any correlation in the neighborhoods and concentration of Chinese restaurants. We chose the **kmeans clustering** method as it is the most common and best suited for our purposes. We also use the Elbow Method and determined the best number of clusters is **5**.

Next, we do a **left join** on the two main dataframe we have: the most common venue dataframe and our original dataframe containing the neighborhood, borough, postal code, latitude and longitude. We also visualize the results of our machine learning by utilizing folium to superimpose the clusters onto a map of London. Lastly, we examine each cluster by slicing the dataframe by cluster.

4. Results

Our final results from the analysis is the following dataframe containing the columns Neighborhood, Borough, PostCode, Latitude, Longitude, Cluster Labels, 1st Most Common Venue to 10th Most Common Venue.

	Neighborhood	Borough	PostCode	Latitude	Longitude	Cluster Labels	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
0	Bayswater	Westminster	W2	51.5156	-0.188645	3	Café	Hotel	Coffee Shop	Pub	Pizza Place	Gym / Fitness Center	Garden	Persian Restaurant
1	Bloomsbury	Camden	WC1	51.5236	-0.122300	3	Garden	Pub	Coffee Shop	Bookstore	Park	Café	Fish & Chips Shop	Plaza
2	Bow	Tower Hamlets	E3	51.5250	-0.026571	2	Pub	Bus Stop	Chinese Restaurant	Park	Breakfast Spot	Metro Station	Bar	Road
3	Brent Cross	Barnet	NW2	51.5649	-0.223325	0	Chinese Restaurant	Arts & Crafts Store	Supermarket	Warehouse Store	Bus Station	Sporting Goods Shop	Clothing Store	Grocery Store
4	Brent Cross	Barnet	NW4	51.6000	-0.216700	0	Chinese Restaurant	Arts & Crafts Store	Supermarket	Warehouse Store	Bus Station	Sporting Goods Shop	Clothing Store	Grocery Store

Fig 1

We drill down and see here the 10 most common venues for each neighborhood and, highlighted in the green boxes, if Chinese restaurants is one of them.

	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	Bayswater	Café	Hotel	Coffee Shop	Pub	Pizza Place	Gym / Fitness Center	Garden	Persian Restaurant	Greek Restaurant	Chinese Restaurant
9	Bloomsbury	Garden	Pub	Coffee Shop	Bookstore	Park	Café	Fish & Chips Shop	Plaza	Cheese Shop	Chinese Restaurant
10	Bow	Pub	Bus Stop	Chinese Restaurant	Park	Breakfast Spot	Metro Station	Bar	Road	Burger Joint	Coffee Shop
11	Brent Cross	Chinese Restaurant	Arts & Crafts Store	Supermarket	Warehouse Store	Bus Station	Sporting Goods Shop	Clothing Store	Grocery Store	Hardware Store	Ethiopian Restaurant
13	Camberwell	Coffee Shop	Pub	Café	Park	Grocery Store	Gastropub	Middle Eastern Restaurant	Beer Bar	Vietnamese Restaurant	Gym / Fitness Center
16	Church End	Turkish Restaurant	Supermarket	Japanese Restaurant	Pizza Place	Pub	Café	Coffee Shop	Hookah Bar	Juice Bar	Korean Restaurant
17	Colindale	Pub	Outdoor Supply Store	Clothing Store	Asian Restaurant	Café	Park	Chinese Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant
23	Finchley	Supermarket	Turkish Restaurant	Japanese Restaurant	Pizza Place	Coffee Shop	Pub	Café	Rugby Pitch	Korean Restaurant	Sports Club
31	Little Venice	Hotel	Café	Coffee Shop	Pub	Gym / Fitness Center	Garden	Pizza Place	Indian Restaurant	Greek Restaurant	Chinese Restaurant
35	Nunhead	Pub	Bus Stop	Hotel	Bakery	Pharmacy	Deli / Bodega	Restaurant	Chinese Restaurant	Platform	Fried Chicken Joint
39	Rotherhithe	Café	Park	Pub	Sporting Goods Shop	Breakfast Spot	Restaurant	Chinese Restaurant	Furniture / Home Store	Metro Station	Pharmacy

Fig 2

Our folium map with the markers for each cluster was rendered as shown below.

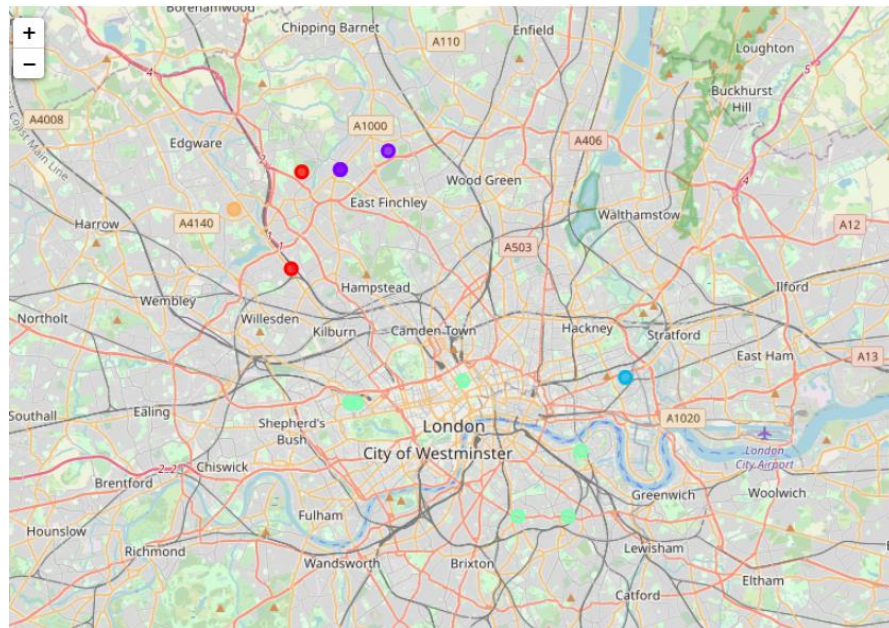


Fig 3

The resulting slicing of the dataframe into each cluster is shown below.

Cluster 1

Neighborhood	Cluster Labels	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	
3	Brent Cross	0	Chinese Restaurant	Arts & Crafts Store	Supermarket	Warehouse Store	Bus Station	Sporting Goods Shop	Clothing Store	Grocery Store	Hardware Store	Ethiopian Restaurant
4	Brent Cross	0	Chinese Restaurant	Arts & Crafts Store	Supermarket	Warehouse Store	Bus Station	Sporting Goods Shop	Clothing Store	Grocery Store	Hardware Store	Ethiopian Restaurant

Fig 4

Cluster 2

	Neighborhood	Cluster Labels	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
6	Church End	1	Turkish Restaurant	Supermarket	Japanese Restaurant	Pizza Place	Pub	Café	Coffee Shop	Hookah Bar	Juice Bar	Korean Restaurant
8	Finchley	1	Supermarket	Turkish Restaurant	Japanese Restaurant	Pizza Place	Coffee Shop	Pub	Café	Rugby Pitch	Korean Restaurant	Sports Club
9	Finchley	1	Supermarket	Turkish Restaurant	Japanese Restaurant	Pizza Place	Coffee Shop	Pub	Café	Rugby Pitch	Korean Restaurant	Sports Club

Fig 5

Cluster 3

Neighborhood	Cluster Labels	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	
2	Bow	2	Pub	Bus Stop	Chinese Restaurant	Park	Breakfast Spot	Metro Station	Bar	Road	Burger Joint	Coffee Shop

Fig 6

Cluster 4

	Neighborhood	Cluster Labels	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	Bayswater	3	Café	Hotel	Coffee Shop	Pub	Pizza Place	Gym / Fitness Center	Garden	Persian Restaurant	Greek Restaurant	Chinese Restaurant
1	Bloomsbury	3	Garden	Pub	Coffee Shop	Bookstore	Park	Café	Fish & Chips Shop	Plaza	Cheese Shop	Chinese Restaurant
5	Camberwell	3	Coffee Shop	Pub	Café	Park	Grocery Store	Gastropub	Middle Eastern Restaurant	Beer Bar	Vietnamese Restaurant	Gym / Fitness Center
10	Little Venice	3	Hotel	Café	Coffee Shop	Pub	Gym / Fitness Center	Garden	Pizza Place	Indian Restaurant	Greek Restaurant	Chinese Restaurant
11	Little Venice	3	Hotel	Café	Coffee Shop	Pub	Gym / Fitness Center	Garden	Pizza Place	Indian Restaurant	Greek Restaurant	Chinese Restaurant
12	Nunhead	3	Pub	Bus Stop	Hotel	Bakery	Pharmacy	Deli / Bodega	Restaurant	Chinese Restaurant	Platform	Fried Chicken Joint
13	Rotherhithe	3	Café	Park	Pub	Sporting Goods Shop	Breakfast Spot	Restaurant	Chinese Restaurant	Furniture / Home Store	Metro Station	Pharmacy

Fig 7

Cluster 5

	Neighborhood	Cluster Labels	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
7	Colindale	4	Pub	Outdoor Supply Store	Clothing Store	Asian Restaurant	Café	Park	Chinese Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant

Fig 8

5. Discussion and Recommendation

Based on an initial look at *Fig 2*, we can see that only 2 neighborhoods – Brent Cross, Barnet and Bow, Tower Hamlets – have Chinese restaurants listed in the top 5 most common venues. Some of the other neighborhoods do not even have Chinese restaurant listed. We now have a clearer picture of which neighborhoods seem to be popular for Chinese restaurants.

From the *Fig 3* map, we can see that the area we want to be focusing on is near the London city center and north of London.

Cluster 2 and 5 (in *Fig 5* and *8*) are removed from our consideration. They do not have Chinese restaurant listed in the 10 most common venues. We are looking for areas where demand for Chinese restaurants is high, thus where Chinese restaurants are a popular locale.

The most viable option here is Cluster 1 and 3. The clustering algorithm has confirmed that the neighborhood Brent Cross and Bow are our best options. However, we must keep in mind that there are other contributing factors that we have not considered due to the lack of public information available such as a breakdown of ethnicities in each neighborhood and market data on Chinese restaurants, i.e., who are more likely to frequent them including a breakdown of age, income, education, ethnicity etc. information. Also, the Census information we obtained was from back in 2011, 10 years ago, therefore the demographic in London may have changed since then.

A city as diverse as London and the 5 selected boroughs with the highest Chinese population only results in 11 neighborhoods that have Chinese restaurants is unlikely.

Given this caveat, our recommendation for where to open a Chinese restaurant is either at **Brent Cross, Barnet or Bow, Tower Hamlet**.

6. Conclusion

In this project, our aim was to identify which neighborhood would be best suited to open a new Chinese restaurant. Our main methodology to determine this was to understand where the demand is for Chinese restaurants. We accomplished this by determining where the concentration of Chinese communities are, then analyzing each neighborhood in those boroughs. The

neighborhoods with the most Chinese restaurants are the ones that were deemed the most viable options as this indicates where the demand is high.

Our recommendation, ultimately, was either Brent Cross, Barnet or Bow, Tower Hamlet. Although in this report, we do acknowledge that the missing data, especially market data on Chinese restaurants is required to give an even more accurate recommendation.

This information would be useful for those looking to open a new Chinese restaurant and wanting to know where the demand for them is high.