

# Capstone Project

## “Wontons in London”

### 1. The Challenge

A restaurant owner recently approached me with his idea to open a Chinese themed restaurant "Wontons" in London, England. London is known for its diversity and range of cultures with more than 300 languages spoken in the city. As a result of this diversity London already offers a great variety of cuisine.

Of course, as a businessman he would like this idea to be as profitable as possible. So, through data science, he has asked me to try and help him decide where in London his restaurant should be located. His understanding of data is limited so my findings should be as palatable and easy to understand as possible. In order for Wontons, or any restaurant, to be successful in such a competitive market like London we must consider the following questions;

- Will it get enough customers?
- Is the location affordable?
- Will it have competition?

Through careful analysis hopefully we can find a location that answers these questions and will maximise profits and minimise costs.

### 2. The Data

For this task I will need to utilise the following data resources;

- **Wikipedia**

Areas of London: [https://en.wikipedia.org/wiki/List\\_of\\_areas\\_of\\_London](https://en.wikipedia.org/wiki/List_of_areas_of_London)

*This data will help us segment the city and is vital for clustering analysis.*

Ethnic groups in London: [https://en.wikipedia.org/wiki/Ethnic\\_groups\\_in\\_London](https://en.wikipedia.org/wiki/Ethnic_groups_in_London)

*This data will help us understand the ethnic population of London and will hopefully give us a good idea of where the restaurant will gain the most custom.*

- **Property Data**

London Property Prices: <https://propertydata.co.uk/cities/london>

*This data will help us understand where the most, and least expensive, locations are to open a restaurant in London.*

- **Foursquare**

Location data about London's existing restaurants and food venues will be used to help us understand how competitive the market is.

### 3. The Methodology

I will break my analysis down into the following three clear steps;

#### 1) Acquire London Property Pricing Data

From my research I have found two useful sites that contain free datasets that will allow me to construct a list of the average property prices in each London borough. The sources can be seen below;

[https://en.wikipedia.org/wiki/List\\_of\\_areas\\_of\\_London](https://en.wikipedia.org/wiki/List_of_areas_of_London)  
<https://propertydata.co.uk/cities/london>

Through careful data scraping and cleaning I can merge the two tables so that we have one table containing "London Borough" and "Average Price". The results will then be made easier to visualise in a chart.

Using this data the client will be given an early indication about how much they will have to spend to acquire a restaurant location and which boroughs will be cheapest to open the restaurant in.

#### 2) Acquire and Explore London Population Data

The second step of my analysis will contain the bulk of the work. Working under the assumption that there is a positive correlation between the size of a Chinese community and the success of a Chinese restaurant, I will attempt to establish where in London the largest Chinese population is.

To acquire this data I will use the following source;

[https://en.wikipedia.org/wiki/Ethnic\\_groups\\_in\\_London](https://en.wikipedia.org/wiki/Ethnic_groups_in_London)

After scraping and cleaning the data I will add coordinates to each borough I will use the Folium package to visualise them on a London map. Then using Foursquare I will explore each borough, looking for their popular venues to analyse and better understand each community. Using this data I will cluster the boroughs based on similarity and examine the results to see how they help the client.

Following this I will merge the data resulting data from steps 1 and 2 and create a scatter plot of "Average Price" vs "Chinese Population" and examine the results.

#### 3) Suggest Locations

Finally, having merged and visualised the results from the previous steps I believe that I will have enough data to suggest one or more locations where the client should look to open his restaurant. Using Foursquare's "Venue" feature I will be able to examine each borough to find local restaurant competition and hopefully from that information make a final decision.

## 4. The Results

In this section I will display the results of each section, commenting on my analysis and observations.

### 1) Acquire London Property Pricing Data

After scraping the Wikipedia page for the list of “Areas of London” and cleaning the messy data I had a table that looked like the below;

	Location	London borough	Post town	Postcode district	Dial Code	OS grid ref
0	Abbey Wood	Bexley	LONDON	SE2	020	TQ465785
1	Acton	Ealing	LONDON	W3	020	TQ205805
2	Addington	Croydon	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY	DA5	020	TQ478728

Doing the same for the “Property Data” gave the following;

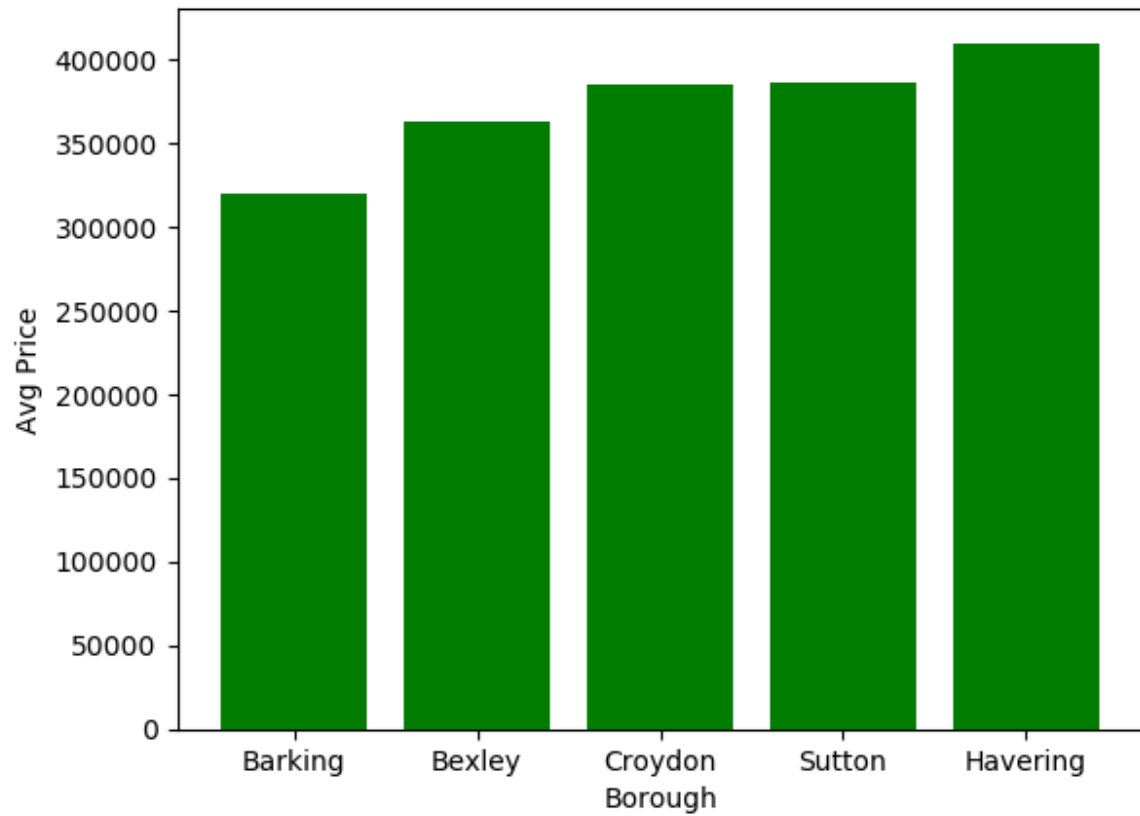
	Area	Avg yield	Avg price	£/sqft	5yr +/-	Explore data
0	BR1	3.3%	£431,123	£461	+23%	Explore data
1	BR2	3.5%	£489,265	£467	+24%	Explore data
2	BR3	3.6%	£440,455	£489	+24%	Explore data
3	BR4	-	£628,666	£451	+21%	Explore data
4	BR5	2.9%	£450,523	£423	+23%	Explore data

Now all that remained to do was to merge the two, leaving us with

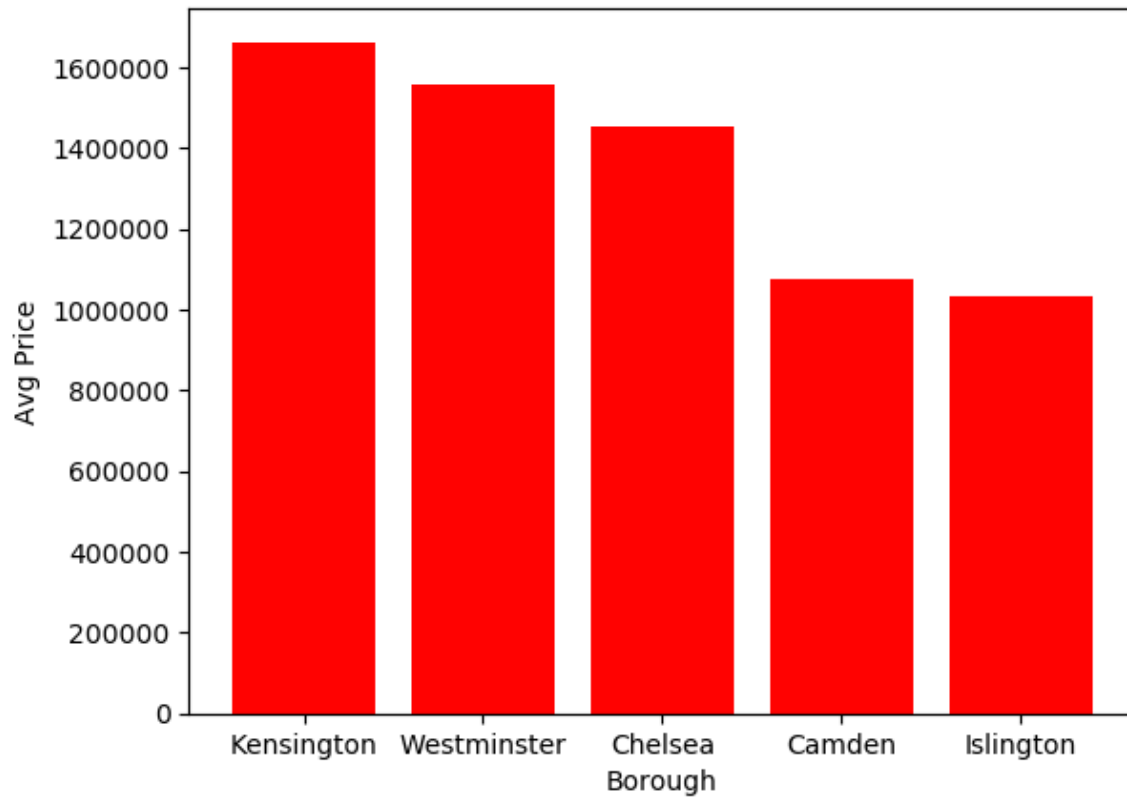
	Location	London borough	Post town	Postcode district	Dial Code	OS grid ref	Avgprice
120	Blackfriars	City	LONDON	EC4	020	TQ318808	£1,016,190
121	Temple	City	LONDON	EC4	020	TQ311809	£1,016,190
134	Bloomsbury	Camden	LONDON	WC1	020	TQ305825	£1,031,839
135	Holborn	Camden	LONDON	WC1	020	TQ305815	£1,031,839
136	King's Cross	Camden and Islington	LONDON	WC1	020	TQ315835	£1,031,839

Through grouping the boroughs and sorting by “average price” we can visualise the cheapest and most expensive areas by looking at the following bar charts.

Least Expensive London Boroughs



Most Expensive London Boroughs



Intuitively we can see that the cheapest areas to buy properties in London are Barking, Bexley and Croydon, respectively, while the most expensive are Kensington, Westminster and Chelsea. If the client were to work off this information alone we would recommend one of the boroughs in the first chart. Of course, however, more information is needed.

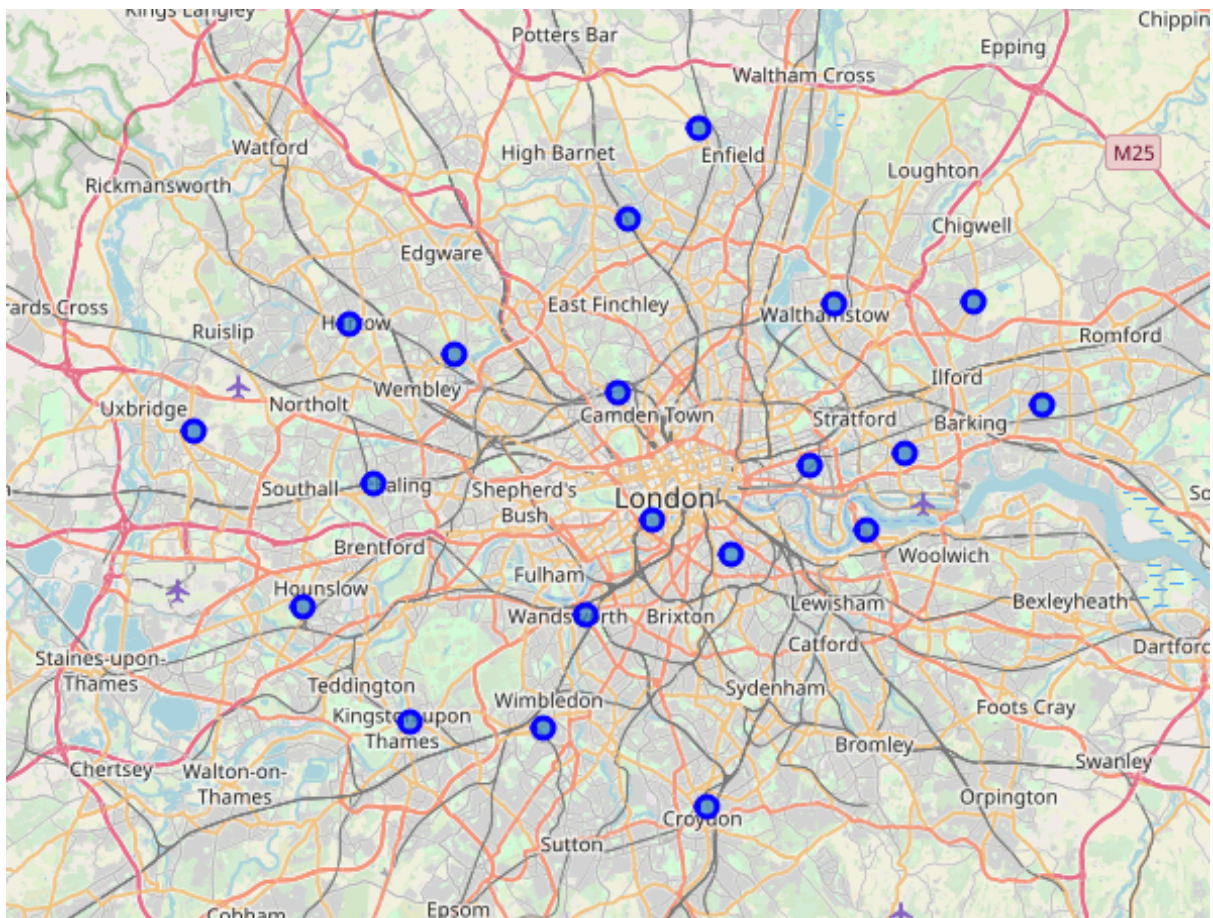
## 2) Acquire and Explore London Population Data

After scraping the Wikipedia page for the list of “Asian population of London” and cleaning the messy data I had a table that looked like the below;

	Rank	London Borough	Indian Population	Pakistani Population	Bangladeshi Population	Chinese Population	Other Asian Population	Total Asian Population
0	1	Newham	42484	30307	37262	3930	19912	133895
1	2	Redbridge	45660	31051	16011	3000	20781	116503
2	3	Brent	58017	14381	1749	3250	28589	105986
3	4	Tower Hamlets	6787	2442	81377	8109	5786	104501
4	5	Harrow	63051	7797	1378	2629	26953	101808

Giving each borough coordinates allowed me to use the Folium python package to visualise the boroughs on a map;

### London Borough Map



Using the Foursquare API service and the “one hot encoding” method I was able to obtain information for the most popular venues of each borough;

```

----Barking and Dagenham----
      venue  freq
0  Construction & Landscaping  0.33
1                Lake         0.33
2                Park         0.33
3      American Restaurant    0.00
4                Museum    0.00

```

```

----Barnet----
      venue  freq
0  Rental Car Location  0.25
1            Bus Stop   0.25
2              Café     0.25
3      Home Service     0.25
4  American Restaurant  0.00

```

And visualise the results in a table as below;

	Borough	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	Barking and Dagenham	Construction & Landscaping	Lake	Park	Women's Store	Fish & Chips Shop	Department Store	Dessert Shop
1	Barnet	Bus Stop	Home Service	Café	Rental Car Location	Women's Store	Electronics Store	Fast Food Restaurant
2	Brent	Middle Eastern Restaurant	IT Services	Bus Station	Fast Food Restaurant	Food Truck	Supermarket	Café
3	Camden	Café	Gym / Fitness Center	Coffee Shop	Gastropub	Pizza Place	Bakery	Museum

Interestingly, out of the 19 boroughs that I was able to gather this information from only two (Hillington and Greenwich) had a Chinese restaurant in their top 5 most common venue. Perhaps indicating a lack of quality offering in the areas.

Using this information I used Machine Learning (K-Means Clustering,) to cluster the boroughs into 6 similar groups. Merging the tables for “Common Venues” and “Asian Populations” allowed me to visualise the results as below;

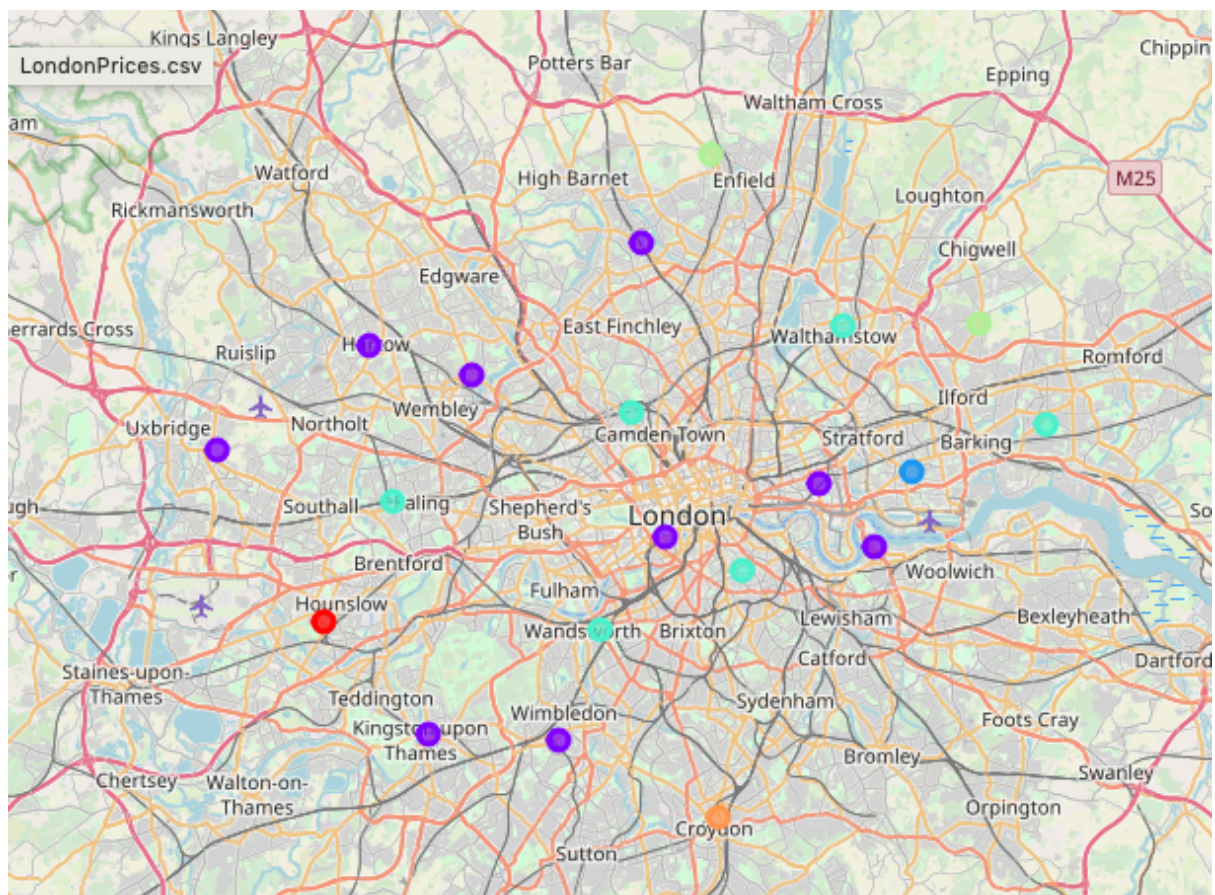


	Rank	London Borough	Indian Population	Pakistani Population	Bangladeshi Population	Chinese Population	Other Asian Population	Total Asian Population	Borough	1st Most Common Venue	2nd Most Common Venue
0	1	Newham	42484	30307	37262	3930	19912	133895	Newham	Café	Grocer Store
1	2	Redbridge	45660	31051	16011	3000	20781	116503	Redbridge	Indian Restaurant	Superm
2	3	Brent	58017	14381	1749	3250	28589	105986	Brent	Middle Eastern Restaurant	IT Serv
3	4	Tower Hamlets	6787	2442	81377	8109	5786	104501	Tower Hamlets	Pizza Place	Canal I

(Where unseen is a column for “Cluster Labels”)

From this table we can visualise the clusters as below;

London Borough Cluster Map



(Where each colour represents a different cluster)

Interesting, the largest average amount of Chinese population can be found in the data's largest cluster, cluster 2 (purple on the map), which is somewhat intuitive knowing that they are clustered on similarity of

behaviour. However, of the 9 boroughs in this cluster only two (again Hillington and Greenwich) have a Chinese restaurant in their top 5 venues. We could say, therefore that the other 7 remaining venues;

- Barnet
- Tower Hamlets
- Westminster
- Brent
- Kingston Upon Thames
- Harrow
- Merton

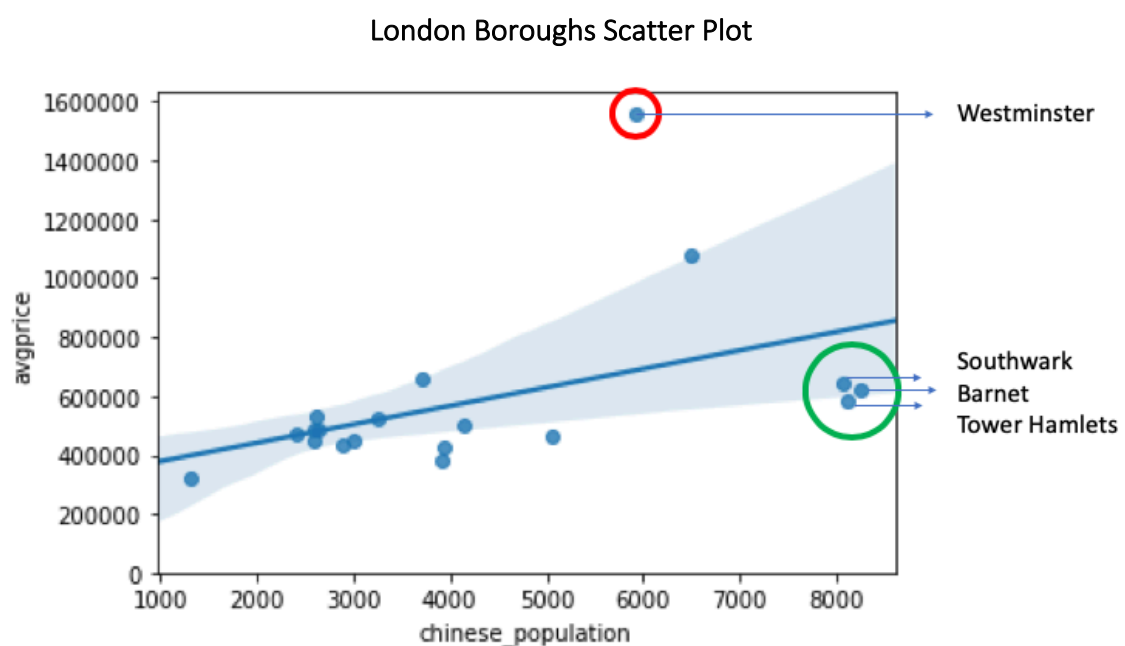
Are prime locations for a quality Chinese restaurant. However, 7 boroughs is too broad, more information is needed.

Through merging the results of the data in the previous step and this the following table (with shape 19,23) is given;

	<b>london_borough</b>	<b>avgprice</b>	<b>rank</b>	<b>indian_population</b>	<b>pakistani_population</b>	<b>bangladeshi_population</b>	<b>chinese_population</b>
<b>0</b>	Barking and Dagenham	319211	18	7436	8007	7701	1315
<b>1</b>	Croydon	384420	10	24660	10865	2570	3925
<b>2</b>	Hounslow	474877	7	48161	13676	2189	2405
<b>3</b>	Enfield	485936	14	11648	2594	5599	2588
<b>4</b>	Ealing	504298	6	48240	14711	1786	4132
<b>5</b>	Redbridge	446953	2	45660	31051	16011	3000
<b>6</b>	Hillingdon	435438	8	36795	9200	2639	2889

(Where unseen are the columns for clusters and most popular venues)

Using this data we can create a scatter plot for “Average Price” vs “Chinese Population”.





From this plot we can clearly see three boroughs that stand out, the boroughs of Barnet, Tower Hamlets and Southwark have the three highest Chinese populations respectively. We can also see that comparative to their populations their property prices are relatively low. It would be my recommendation that, based on the initial assumption that these would be the three best possible locations for a Chinese restaurant. We can also note that Barnet and Tower Hamlets were clustered together. Interestingly we can see from the regression line that there is a positive correlation between average property price and size of Chinese population, however I don't believe that this is relevant to the question at hand. Westminster is a clear outlier, but due to its high average price it is not important to this study.

### 3) Suggest Locations

Now we have three potential boroughs;

- Barnet
- Tower Hamlets
- Southwark

We can use Foursquare API's "venue" feature to examine each further and decide which location should be suggested to the client.

We can do this simply by searching each borough with radius 5000 meters to determine how many existing Chinese restaurants are in each area. Repeating for each gives us a table and shape that looks as below;

(22, 17)

	id	name	categories	referralId	hasPerk	location.lat	location.lng
0	4f8db193e4b05cdaa118f17b	Chinese	[{"id": "4bf58dd8d48988d145941735", "name": "C..."}]	v-1578334510	False	51.616079	-0.132721
1	516d4b05e4b0aced3b68e637	Chinese Golden Buffet	[{"id": "4bf58dd8d48988d145941735", "name": "C..."}]	v-1578334510	False	51.632861	-0.126294
2	4c5b0bfa15aaef3b4a19c580	Honeymoon Chinese Restaurant	[{"id": "4bf58dd8d48988d145941735", "name": "C..."}]	v-1578334510	False	51.632137	-0.128725
3	4bd4a67f6f649521bdd36dec	Wood Green Chinese Restaurant	[{"id": "4bf58dd8d48988d145941735", "name": "C..."}]	v-1578334510	False	51.614642	-0.128742
4	51f2d55a498e87240a32f1bd	Chinese Restaurant	[{"id": "4bf58dd8d48988d145941735", "name": "C..."}]	v-1578334510	False	51.595484	-0.112953

Where the results are;

Barnet: 22 existing competitors,  
Tower Hamlets; 50 existing competitors,  
Southwark; 50 existing competitors.

## 5. Discussion

Clearly, we can see that Barnet has the least amount of competition of the three potential locations. Of the entire dataset Barnet also has the highest number of Chinese population (8259) and has a relatively moderate average property price (£618,697). Given this, I can conclude that Barnet would make the most sense as the location for a new Chinese restaurant, “Wontons”, in London.

## 6. Conclusion

I am relatively happy with the report considering that it has delivered, what I consider to be, a clear and obvious location, which was the aim of the process. However, I will concede that it is not perfect. Out of the 32 boroughs of London, due to the limited datasets, we only ended up with 19 in our final merged table. Also, the dataset for the Asian population of London is taken from a census from 2011 and thus may not completely relevant at this time. There is also no guarantee that the client will be able to find an available property in Barnet suitable for a restaurant and is that of the average price. Finally, and perhaps most importantly, for a restaurant to be successful the food must be good!

If the data was available I believe this process could be replicated, and would be effective, for any city, and any cuisine, around the world.