

IBM Data Science Capstone

Restaurant Location Cluster in London

Introduction

An important step in opening a business is zeroing on its location; the revenue of a business is highly correlated to how good its location is - businesses located highly close to each other have to split potential revenue while businesses located very remotely suffer in the other direction. So it would be essential to identify potential areas that would be promising in terms of revenue it could generate from customers.

Clustering of Businesses

It is generally observed that businesses open shop close-by, contrary to general reasoning that it could hurt customer turnout. Such clustering phenomenon though, is explained by Game Theory when you have commodities which are reasonably interchangeable, such as retailers.

Consider a retailer/business that has a monopoly. It could simply open shop at its location of choice and force customers to travel for them. However, the threat of competition forces stores to try to determine a location that captures maximum market share. In other words, they want to be in central location for their target audience, minimizing the distance consumers have to travel. In the mean time, all competing businesses are simultaneously making the same decision which means the resulting stores end up clustered together. If a retailer opens a new location away from the current clustering, there are two potential results:

1. It will fail to capture enough consumers and eventually close.
2. It will become successful causing competitive stores to locate nearby.

Either way, clustering remains the norm. Such a behaviour is sometimes explained in economics using Nash Equilibrium.

Problem & Target Audience

Identify restaurants according to cuisines with most footfall in the city, in our case, London and cluster the neighbourhoods accordingly to identify where one could open similar restaurants.

Restaurants wanting to open up in areas where demand is high could make use of the conclusions from the report. If the business had an approach to avoid areas where there are already serving large number of customers and instead focus on areas outside the high footfall areas may instead avoid these areas when considering to open a new business.

Approach

- Identify restaurant cuisines and cluster them according to neighbourhoods(Boroughs) in London. This will identify the cuisine most famous in the neighbourhood.
- Identify the type of customers one can expect using land prices as a proxy. If there are multiple cuisines available per borough, businesses might want to focus on the type of customers it might want to attract. The price of land prices is used as a proxy to determine the average resident wealth.

Data

The data for neighbourhoods in London can be scrapped from their wikipedia page. The data includes, amongst others: Borough, Local authority, Political control, Area (s.q. mi), Population (2013 est.), Co-ordinates. For the sake of analysis, only *coordinates* and *boroughs* are necessary. Those coordinates will be later used to build a list of business within a certain radius from it and fed into the Foursquare API to get the details of number so customers visiting them. The frequency of visits will be

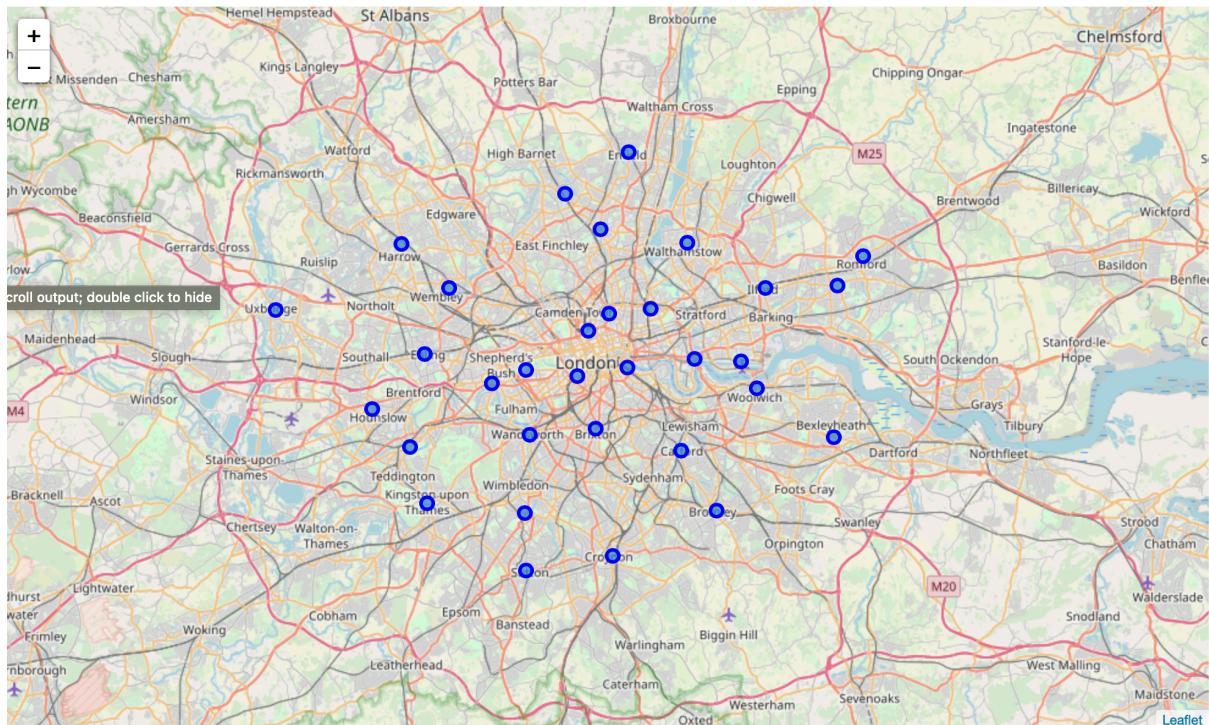
calculated and the boroughs will be clustered according to clustering algorithms to identify the boroughs with high demand for certain businesses.

Land Prices for the city of London can be obtained from the *HM Land Registry* website Land value estimates for policy appraisal 2017 (The sheet one containing the residential land prices has been considered). The data after extraction and cleaning:

Neighbourhood	Latitude	Longitude	Price (Pounds/ha)
Barking and Dagenham	51.5607	0.1557	5400000
Barnet	51.6252	-0.1517	24900000
Bexley	51.4549	0.1505	10300000
Brent	51.5588	-0.2817	16800000
Bromley	51.4039	0.0198	17700000
Camden	51.5290	-0.1255	45700000
Croydon	51.3714	-0.0977	18300000
Ealing	51.5130	-0.3089	27200000
Enfield	51.6538	-0.0799	30200000
Greenwich	51.4892	0.0648	21700000
Hackney	51.5450	-0.0553	35850000
Hammersmith and Fulham	51.4927	-0.2339	75000000
Haringey	51.6000	-0.1119	43100000
Harrow	51.5898	-0.3346	12200000
Havering	51.5812	0.1837	8400000
Hillingdon	51.5441	-0.4760	16600000
Hounslow	51.4746	-0.3680	11800000
Islington	51.5416	-0.1022	39500000
Kensington and Chelsea	51.5020	-0.1947	180600000
Kingston upon Thames	51.4085	-0.3064	18500000
Lambeth	51.4607	-0.1163	24600000
Lewisham	51.4452	-0.0209	41700000
Merton	51.4014	-0.1958	19100000
Newham	51.5077	0.0469	18300000
Redbridge	51.5590	0.0741	6300000
Richmond upon Thames	51.4479	-0.3260	28000000
Southwark	51.5035	-0.0804	67300000
Sutton	51.3618	-0.1945	10000000

Tower Hamlets	51.5099	-0.0059	29800000
Waltham Forest	51.5908	-0.0134	14600000
Wandsworth	51.4567	-0.1910	42500000
Westminster	51.4973	-0.1372	113300000

Neighbourhoods identified using the above table using folium package.



Methodology Results

Using the neighbourhood details (name, co-ordinates) in the four square API, we can obtain the venues (in this case, 100) in the range (lets say, 500 metre) of the borough co-ordinates and obtain details of the number of people who visit those places. The type and variety of venues are also provided.

1. Exploratory Data Analysis:

- We can concentrate only on restaurants and remove other venues like museums, parks, etc.
- Also casual places like pubs, coffee shops are neglected; they are too common to provide any valuable insight.

- The restaurants are clubbed according to their cuisine; the choice may seem arbitrary but they are broadly grouped according to their regions except for cuisines which stand out from their region: for e.g. Italian and French restaurants are not clubbed inside Mediterranean Restaurants due to their special preference amongst patrons.
- The borough “Barking and Dagenham” is neglected because the API doesn’t refer to restaurants of interest in that area.

Neighbourhood	Latitude	Venue	Venue Long.	Venue Category
Barnet	51.6252	The Atrium	51.624726	Café
Barnet	51.6252	Beaconsfield Road (BF)	51.622827	Bus Stop
Barnet	51.6252	Oakleigh Cafe	51.623412	Café
Bexley	51.4549	Wilko	51.456257	Furniture / Home Store
Bexley	51.4549	Zizzi	51.455929	Italian Restaurant

The categories of restaurants/places are clubbed as under:

South Asian = ['Chaat', 'Pakistani', 'Indian']

Chinese = ['Asian', 'Chinese', 'Szechuan', 'Cantonese', 'Dim Sum', 'Dumpling']

East Asian = ['Okonomiyaki', 'Japanese', 'Ramen', 'Sushi', 'Korean']

South East Asian = ['Malay', 'Thai', 'Vietnamese']

Mediterranean = ['Mediterranean', 'Greek', 'Spanish', 'Portuguese', 'Tapas']

Middle Eastern = ['Falafel', 'Turkish', 'Persian', 'Middle', 'Kebab', 'Lebanese']

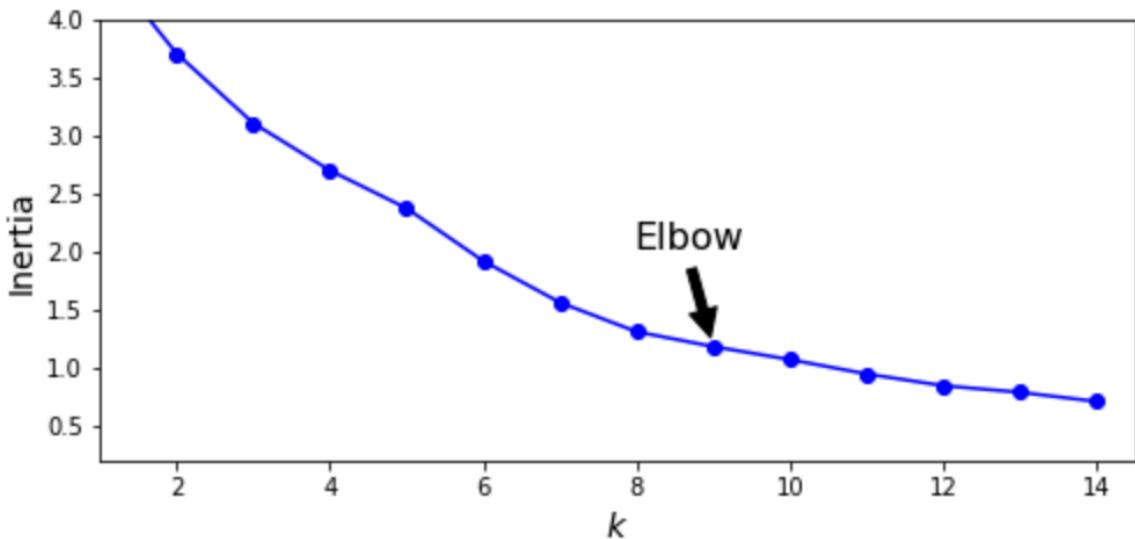
European = ['Scandinavian', 'Polish', 'English', 'European', 'German']

Latin American = ['South', 'Argentinian', 'Latin', 'Brazilian', 'Mexican', 'Cajun']

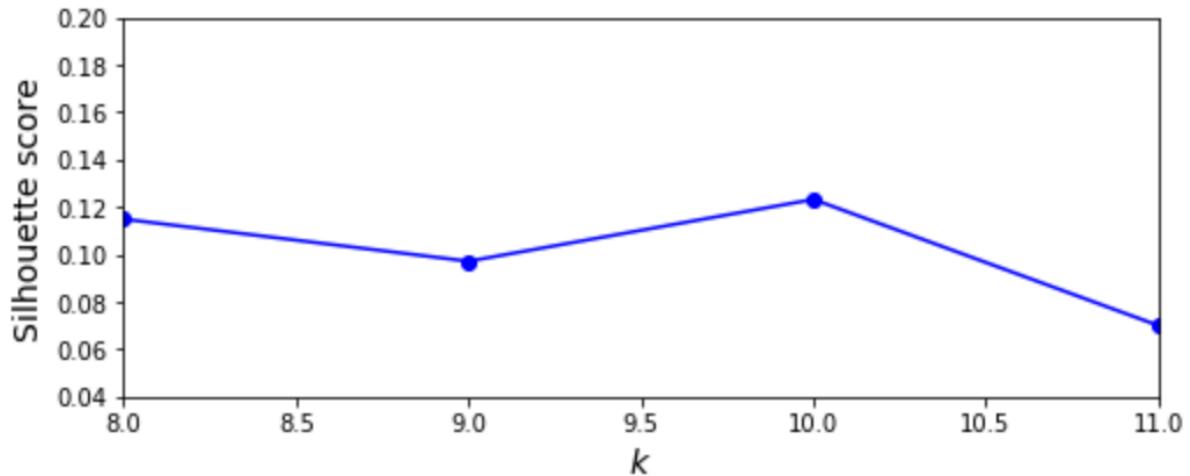
- Grouping neighbourhoods and finding the mean of the frequency of each category and sort the venues.

2. Algorithm

- K-Means is used to cluster neighbourhoods. The number of cluster is determined using the elbow method.

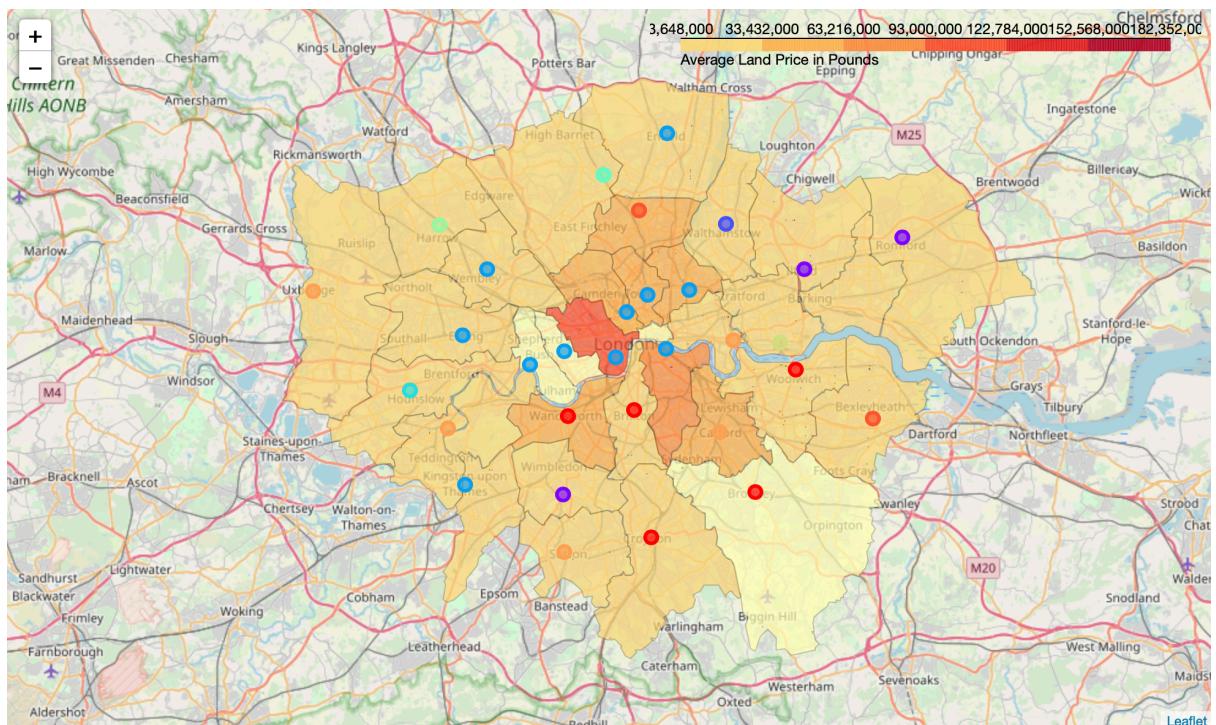


The cluster at which the rate of decline in inertia falls flat is considered the reasonable number of clusters. Due to ambiguity in selection: 8-10 clusters are considered for silhouette-score analysis and 10 clusters are chosen. The peaking cluster of the score is taken.



Results

London boroughs are divided into 10 clusters and the most frequently visited restaurant cuisines are shown below. Choropleth map of land prices are layered to compared neighbourhoods.



Discussion

- Consider the blue cluster popups on the map. The similar points are spread across neighbourhoods of varying land prices.
 - Italian Restaurants and fast food joints seem to be the most visited amongst the selected.
 - Cafe are universal and not a surprising inclusion as the most visited places. So we may ignore the areas with cafes and consider the next most visited if one were to find the clustered area.
 - If a potential restaurant wants to open a restaurant to cater to niche customers, it could prefer to open in a ‘redder’ borough.
 - The below shows the most frequently visited restaurant cuisines by the boroughs.
 - If one were to open Italian Restaurant, clustered here in cluster=3, they have to choose between one of Hillingdon and Richmond Upon Thames. The choropleth map underneath shows the Richmond is pricier. Thus, richer customers can be targeted here for an Italian Restaurant.

Neighbourhood	1st	2nd	Cluster Labels
---------------	-----	-----	----------------

Lambeth	Latin American Restaurant	Caribbean Restaurant	0.0
Bromley	Burger Joint	Mediterranean Restaurant	0.0
Croydon	Mediterranean Restaurant	South East Asia Restaurant	0.0
Wandsworth	Pizza Place	Chinese Restaurant	0.0
Greenwich	Fast Food Restaurant	Sandwich Place	0.0
Merton	Fast Food Restaurant	Café	1.0
Havering	Café	Fast Food Restaurant	1.0
Redbridge	Fast Food Restaurant	Sandwich Place	1.0
Waltham Forest	Vegetarian / Vegan Restaurant	Pizza Place	2.0
Kensington and Chelsea	Continental Restaurant	Café	3.0
Islington	Mediterranean Restaurant	Middle Eastern Restaurant	3.0
Westminster	Sandwich Place	East Asian Restaurant	3.0
Hammersmith and Fulham	Café	Italian Restaurant	3.0
Kingston upon Thames	Café	Italian Restaurant	3.0
Enfield	Café	Sandwich Place	3.0
Ealing	Mediterranean Restaurant	South East Asia Restaurant	3.0
Camden	Café	Mediterranean Restaurant	3.0
Brent	American Restaurant	South Asia Restaurant	3.0
Southwark	Restaurant	French Restaurant	3.0
Hackney	Café	Middle Eastern Restaurant	3.0
Hounslow	Café	Chinese Restaurant	4.0
Barnet	Café	Vegetarian / Vegan Restaurant	5.0
Harrow	South East Asia Restaurant	South Asia Restaurant	6.0
Newham	Sandwich Place	Chinese Restaurant	7.0
Hillingdon	Italian Restaurant	Pizza Place	8.0
Richmond upon Thames	Italian Restaurant	South Asia Restaurant	8.0
Sutton	Pizza Place	Sandwich Place	8.0
Tower Hamlets	Italian Restaurant	Café	8.0

Lewisham	Italian Restaurant	Mediterranean Restaurant	8.0
Haringey	Fast Food Restaurant	Mediterranean Restaurant	9.0
Bexley	Italian Restaurant	Fast Food Restaurant	9.0

Conclusion

The clustering method can be bettered using other algorithms and certain choices of cuisines. But, a basic analysis of the cuisine choice and potential customers can provide a reasonable choice to open a restaurant using the above data and mapping.