

Restaurant Location Cluster in London

IBM Data Science Capstone

Akash M



Photo by [Massimiliano Morosinotto](#) on [Unsplash](#)



Clustering of Business

- Businesses open shop close-by when you have commodities which are reasonably interchangeable, such as retailers.
- Explained by Game Theory using Nash Equilibrium.
- If a retailer opens a new location away from the current clustering :
 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.



Source: <https://www.freepik.com>

Problem and Approach

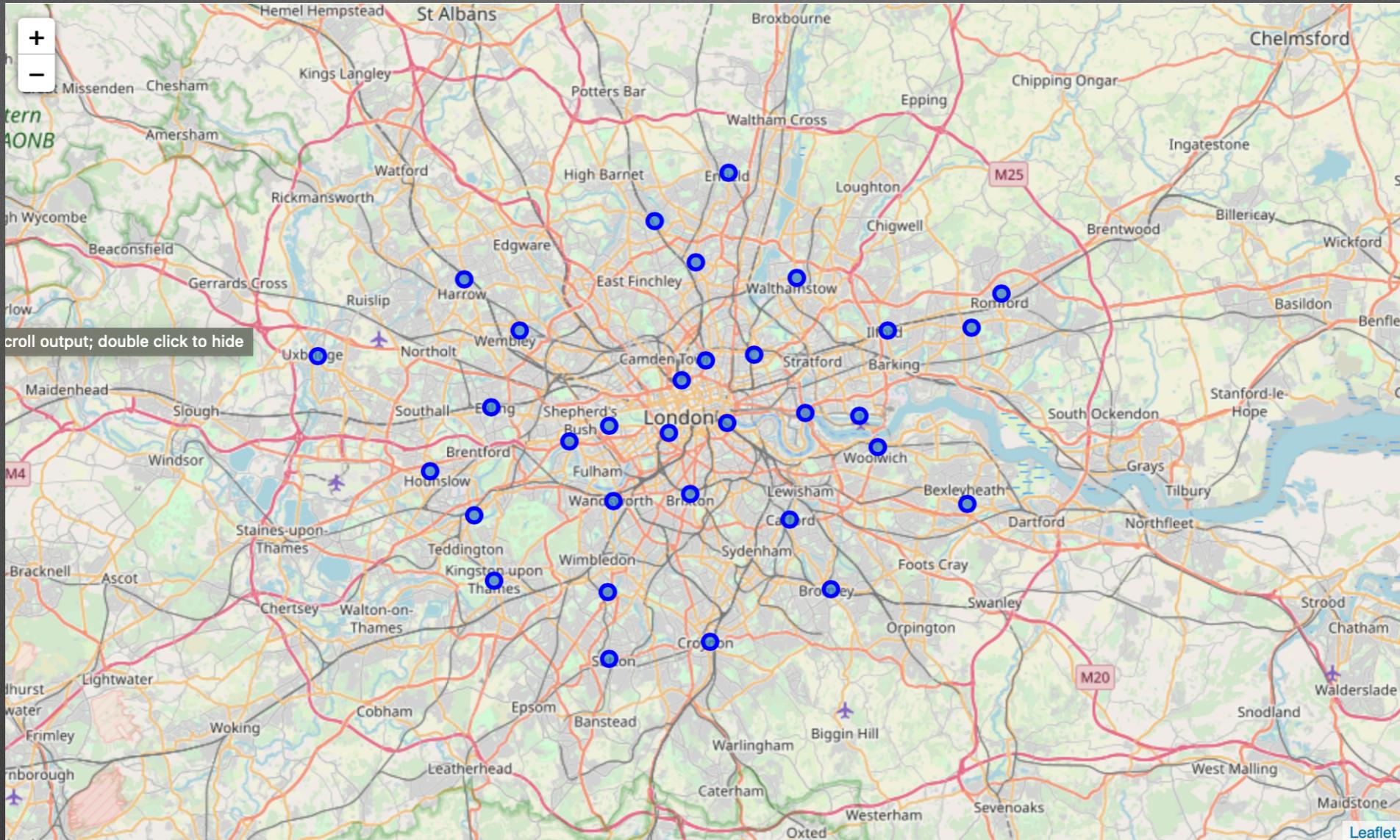
Problem & Target Audience

- Identify restaurants according to cuisines with most footfall in London
- Cluster the neighbourhoods accordingly to identify where the one could open similar restaurants.

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.

Data Sources



- Neighbourhoods (*coordinates* and *boroughs*) in London were scrapped from their wikipedia page.
- List of business within a certain radius from it obtained from Foursquare API.
- 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).

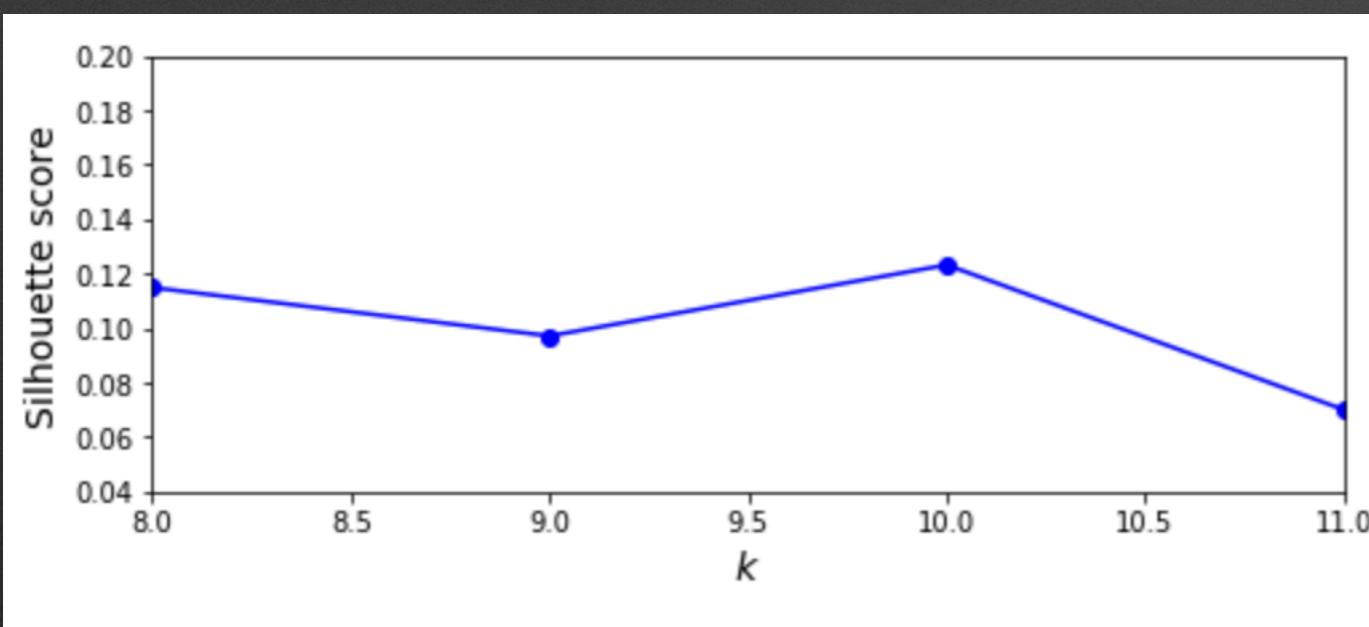
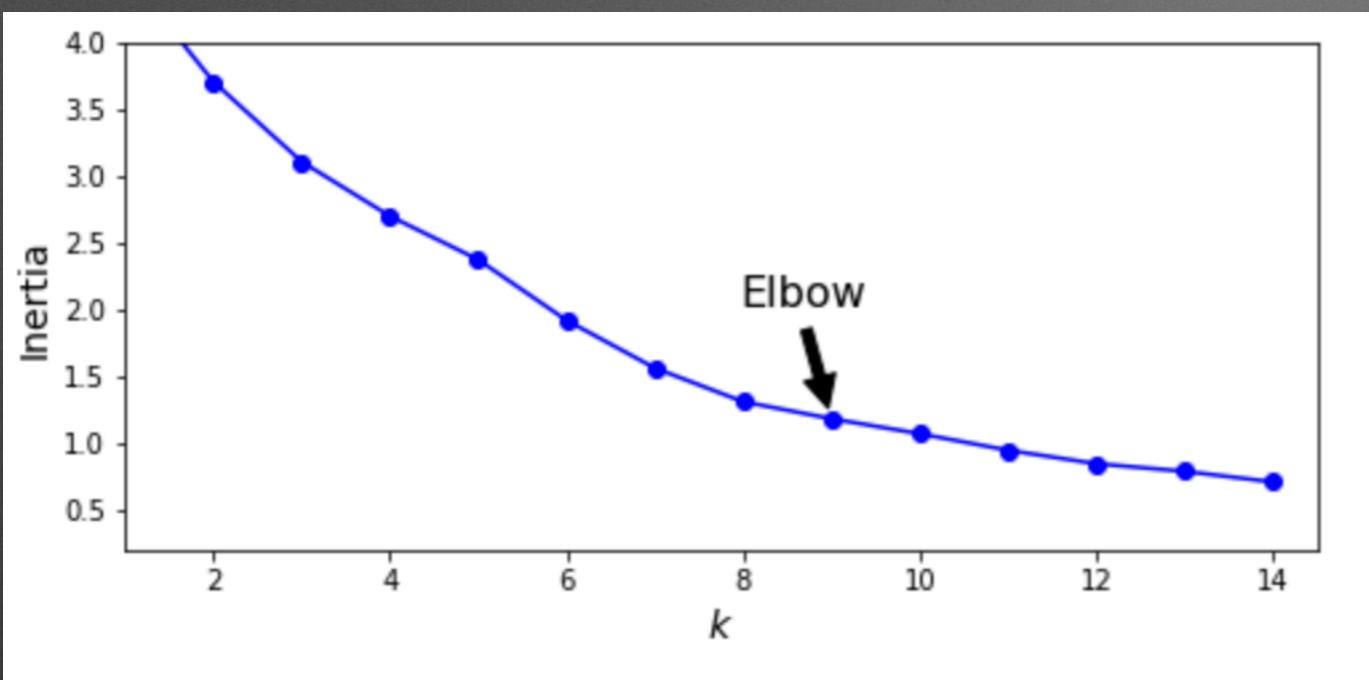
Methodology

- 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; ignore other venues like museums, parks, etc
- The restaurants are clubbed according to their cuisine for example: 'Chaat', 'Pakistani', 'Indian'
- are considered South Asian.
- Grouping neighbourhoods and finding the mean of the frequency of each category and sort the venues.

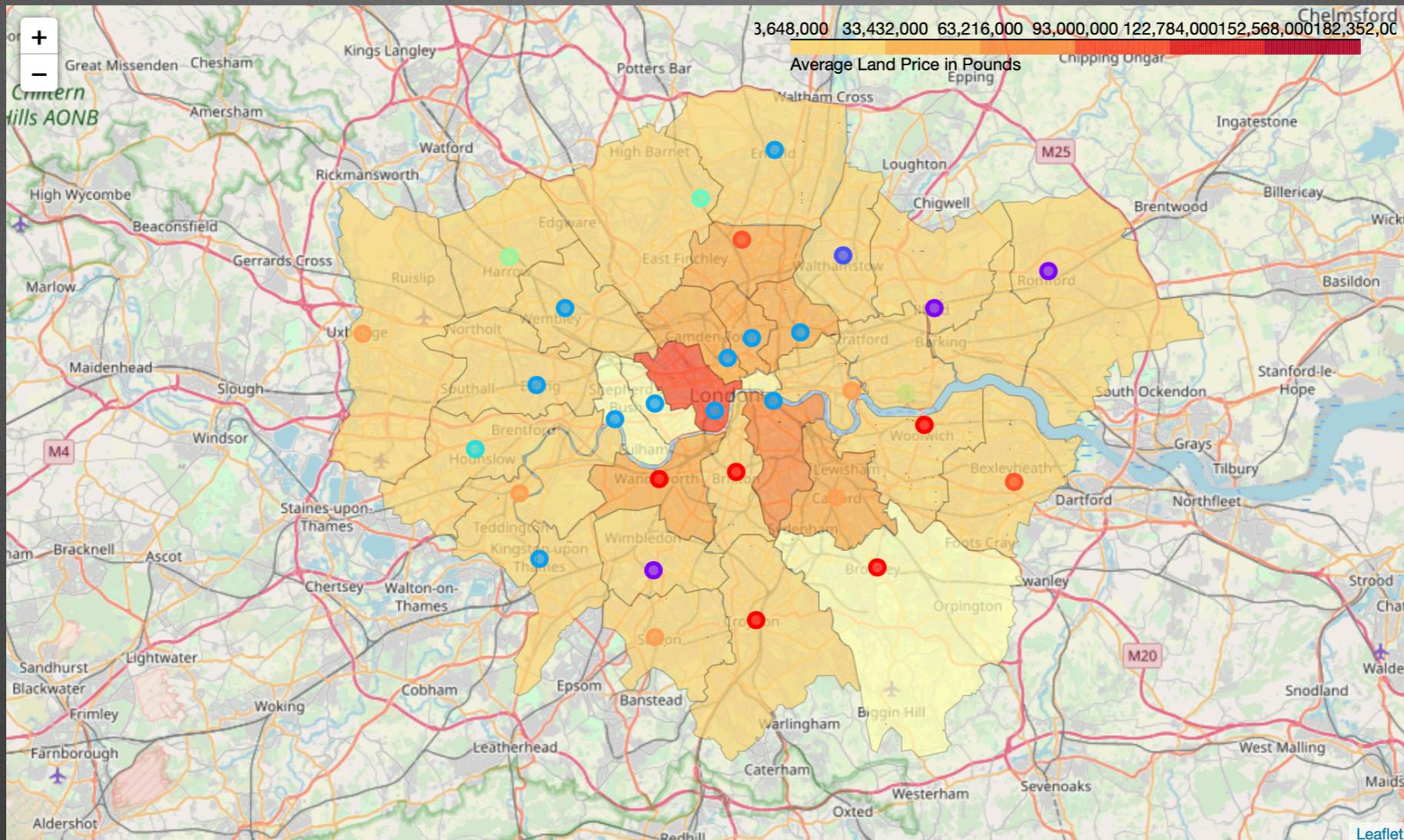
Algorithm

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

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



Results



- Italian Restaurants and fast food joints seem to be the most visited amongst the selected.
- Consider the blue cluster popups on the map. The similar points are spread across neighbourhoods of varying land prices. If a potential restaurant wants to open a restaurant to cater to niche customers, it could prefer to open in a ‘redder’ borough. 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.