Real Estate Affordability based on Price to Income and Venue Data Analysis of London, UK

Elnaz Limouchi

A. Introduction

A.1 Problem Statement

London is considered to be one of the most important global cities around the world and has been called the most powerful, most desirable, most influential, most visited, most expensive, innovative, sustainable, most investment-friendly, and most-popular-for-work city. London exerts a considerable impact upon the arts, commerce, education, entertainment, fashion, finance, healthcare, media, professional services, research and development, tourism and transportation [1].

As mentioned earlier, London is one of the most worldwide expensive cities, the type of business they want to install is less intense. Considering the city residents, it is desirable to choose the regions where real estate are more affordable. At the same time, they may want to select the area according to the social places density. However, it is difficult to obtain information that will guide investors in this direction, nowadays.

Considering all mentioned issues, the problem can be analyzed by creating a map of London and inserting the information on that, and finally clustering each area based on the density of venue.

A.2 Data Description

In order to perform the analysis, these datasets are considered:

- Ratio of House Prices to Earnings (full-time workers by place of work) of London area published by UK Office of National Statistics is used. This dataset shows the average house price/earnings ratio, which is an important indicator of housing affordability. Ratios are calculated by dividing house price by the median earnings of a borough. The Annual Survey of Hours and Earnings (ASHE) is based on a 1 per cent sample of employee jobs. Information on earnings and hours is obtained in confidence from employers. It does not cover the self-employed nor does it cover employees not paid during the reference period. The statistics used are workplace based full-time individual earnings. For this work, I use the data related to 2016 [2].
- In order to calculate the latitude and longitude of different borough, I use geopy.
- I used Forsquare API to get the most common venues of given Borough of London [3].
- I used the Search Nearby service from Google Maps to employ the center coordinates of the each Borough [4].

B. Methodology

In this work, *GitHub* repository is used as the database. The data includes the *Borough (Name)*, *Ratio*, *Latitude* and *Longitude* information of the London area.

	Name	Ratio	Latitude	Longitude
0	City of London	17.51	51.507322	-0.127647
1	Barking and Dagenham	10.55	51.554117	0.150504
2	Barnet	15.19	51.653090	-0.200226
3	Bexley	10.40	51.441679	0.150488
4	Brent	15.93	51.441635	0.234519

Figure 1 Price to Income Ratio Data

Using python **folium** library, geographic details of London and its boroughs are visualized. The map of London with boroughs superimposed on top is created where the latitude and longitude values are utilized to show the map in Fig. 2.

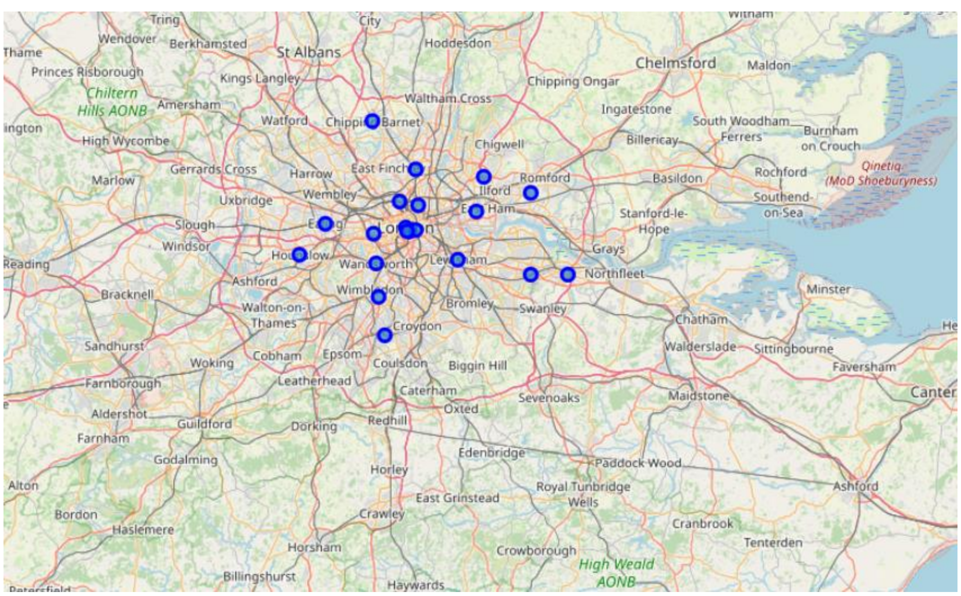

Figure 2 Boroughs of London

The list of venues name, category, and latitude and longitude information are obtained using *Forsquare API.* It is considered to have the limit as 100 venue and the radius as 500 meter for each borough. *Foursquare* returned 100 venues as results which the head of list is shown in Fig. 3.

	name	categories	lat	lng
0	National Gallery	Art Museum	51.508876	-0.128478
1	Trafalgar Square	Plaza	51.507987	-0.128048
2	East Trafalgar Square Fountain	Fountain	51.508088	-0.127700
3	Corinthia Hotel	Hotel	51.506607	-0.124460
4	National Portrait Gallery	Art Gallery	51.509438	-0.128032

Figure 3 List of venues of London by Foursquare

Fig. 4 shows the merged table of boroughs and venues.

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	City of London	51.507322	-0.127647	National Gallery	51.508876	-0.128478	Art Museum
1	City of London	51.507322	-0.127647	Trafalgar Square	51.507987	-0.128048	Plaza
2	City of London	51.507322	-0.127647	East Trafalgar Square Fountain	51.508088	-0.127700	Fountain
3	City of London	51.507322	-0.127647	Corinthia Hotel	51.506607	-0.124460	Hotel
4	City of London	51.507322	-0.127647	National Portrait Gallery	51.509438	-0.128032	Art Gallery

Figure 4 Merged table of boroughs and venues

187 unique categories were returned by *Foursquare*, then I created a table which shows list of top 10 venue category for each borough in below table.

	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Barking and Dagenham	Bus Stop	Convenience Store	Liquor Store	Grocery Store	Yoga Studio	Filipino Restaurant	Fountain	Food Truck	Food & Drink Shop	Food
1	Barnet	Coffee Shop	Park	Pharmacy	Convenience Store	Pub	Grocery Store	Bookstore	Restaurant	Italian Restaurant	Sandwich Place
2	Bexley	Pub	Fast Food Restaurant	Indian Restaurant	Greek Restaurant	Breakfast Spot	Italian Restaurant	Train Station	Toy / Game Store	Chinese Restaurant	Food & Drink Shop
3	Brent	Pizza Place	Pub	Convenience Store	Park	Farm	Food & Drink Shop	Food	Flea Market	Fish Market	Fish & Chips Shop
4	Camden	Pub	Coffee Shop	Burger Joint	Market	Café	Italian Restaurant	Ice Cream Shop	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Music Venue

Figure 5 List of top 10 venue category for each borough

Since there are some common venue categories in boroughs, the unsupervised learning **K-means algorithm** is used to cluster the boroughs. K-Means algorithm is one of the most common cluster method of unsupervised learning.

First, I run K-Means to cluster the boroughs into 2 clusters because when I analyze the K-Means with elbow method it ensured me the 2 degree for optimum k of the K-Means.

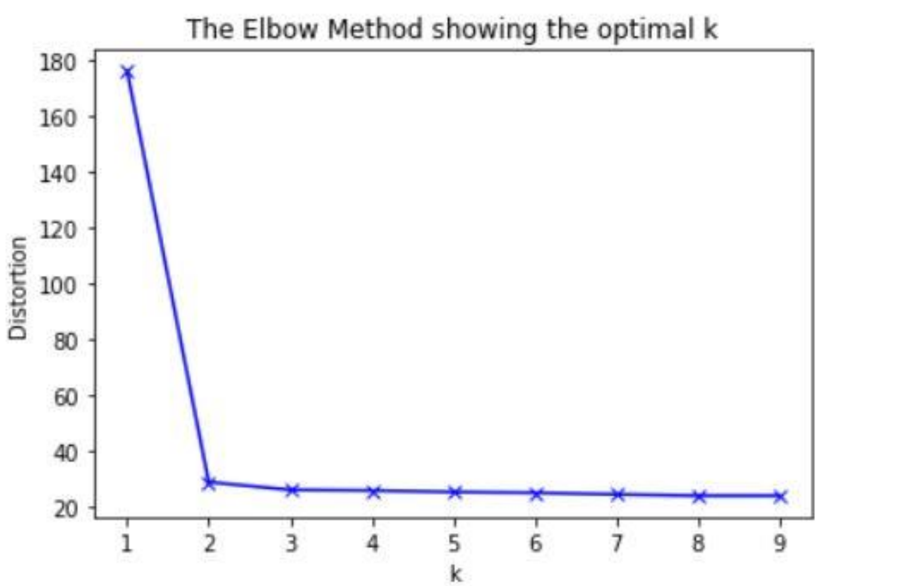

Figure 6 the elbow graph to find the Optimal K

Fig. 7 shows the result.

	Name	Ratio	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	9th Most Common Venue	10th Most Common Venue
0	City of London	17.51	51.507322	-0.127647	0	Theater	Hotel	Japanese Restaurant	Plaza	Art Gallery	Sandwich Place	Garden	Ice Cream Shop	Ramen Restaurant	Steakhouse
1	Barking and Dagenham	10.55	51.554117	0.150504	1	Bus Stop	Convenience Store	Liquor Store	Grocery Store	Yoga Studio	Filipino Restaurant	Fountain	Food Truck	Food & Drink Shop	Food
2	Barnet	15.19	51.653090	-0.200226	1	Coffee Shop	Park	Pharmacy	Convenience Store	Pub	Grocery Store	Bookstore	Restaurant	Italian Restaurant	Sandwich Place
3	Bexley	10.40	51.441679	0.150488	1	Pub	Fast Food Restaurant	Indian Restaurant	Greek Restaurant	Breakfast Spot	Italian Restaurant	Train Station	Toy / Game Store	Chinese Restaurant	Food & Drink Shop
4	Brent	15.93	51.441635	0.234519	1	Pizza Place	Pub	Convenience Store	Park	Farm	Food & Drink Shop	Food	Flea Market	Fish Market	Fish & Chips Shop

Figure 7 Clustering Result

We can also estimate the number of 1st Most Common Venue in each cluster. Thus, we can create a bar chart which may help us to find proper labels for each cluster.

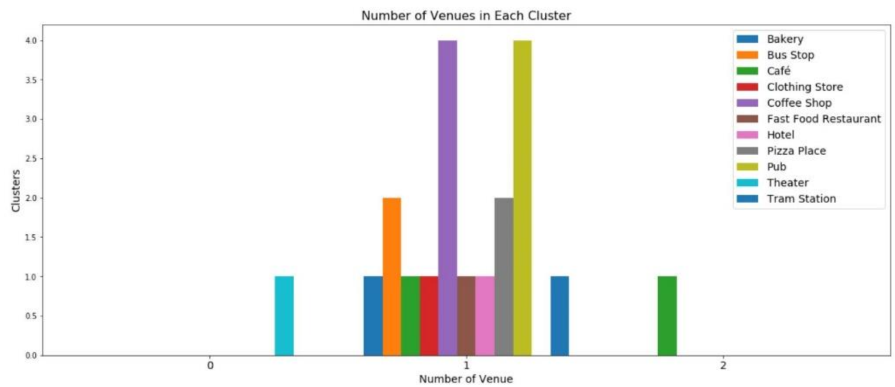

Figure 8 Number of Venues in Each Cluster

One of my aim was also show the number of top 2 venues information for each borough on the map. Thus, I grouped each borough by the number of top 3 venues and I combined those information in *Join* column.

	Borough	Join
0	Barking and Dagenham	2 Bus Stop, 1 Convenience Store, 1 Grocery Store
1	Barnet	4 Coffee Shop, 2 Bookstore, 2 Convenience Store
2	Bexley	2 Fast Food Restaurant, 2 Pub, 1 Breakfast Spot
3	Brent	1 Convenience Store, 1 Park, 1 Pizza Place
4	Camden	9 Pub, 7 Coffee Shop, 5 Burger Joint

Figure 9 Join Column

C. Results

By merging the new variables with the related cluster information in our main master table.

	Name	Ratio	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	9th Most Common Venue	10th Most Common Venue	Borough	Join	Labels
0	City of London	17.51	51.507322	-0.127647	0	Theater	Hotel	Japanese Restaurant	Plaza	Art Gallery	Sandwich Place	Garden	Ice Cream Shop	Ramen Restaurant	Steakhouse	City of London	6 Theater, 5 Hotel, 3 Art Gallery	Cafe Venues
1	Barking and Dagenham	10.55	51.554117	0.150504	1	Bus Stop	Convenience Store	Liquor Store	Grocery Store	Yoga Studio	Filipino Restaurant	Fountain	Food Truck	Food & Drink Shop	Food	Barking and Dagenham	2 Bus Stop, 1 Convenience Store, 1 Grocery Store	Multiple Social Venues
2	Barnet	15.19	51.653090	-0.200226	1	Coffee Shop	Park	Pharmacy	Convenience Store	Pub	Grocery Store	Bookstore	Restaurant	Italian Restaurant	Sandwich Place	Barnet	4 Coffee Shop, 2 Bookstore, 2 Convenience Store	Multiple Social Venues

Figure 10 merging all the results

This work aims to visualize the properties price to income ratio with choropleth style map. Thus, first I downloaded a json file. Finally, I created choropleth map which also has the below information for each borough

- Borough name,

- Cluster name,
- Price to income Ratio,
- Top 2 number of venue

D. Discussion

As mentioned earlier, London is a big city with a high population density. The total number of measurements and population densities of the 20 (in this work) districts in total can vary. As there is such a complexity, very different approaches can be tried in clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same high quality results for this metropole.

Using Kmeans algorithm as part of this clustering study. Based on the Elbow method, the optimum k value is set to 2. However, only 20 district coordinates were used.

Also, data analysis was done through this information by adding the coordinates of districts and affordability as static data on GitHub.

This work finalized by visualizing the data and clustering information on the London map.

As a result, people can achieve better outcomes through their access to the platforms where such information is provided in order to make a decision regarding business start/development.

Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.

References

[1] London, UK

[2] London DataStore

[3] Foursquare

[4] Google Maps