Comparative analysis of London Boroughs using k-means clustering

Aim of study

- London constitutes a home to around 9 million people and is one of the top global financial centres, as well as a popular tourist destination. It is also one of the most expensive European cities to live in according to Mercer 2020 Cost of Living Survey.
- This aim of this study was a more in-depth view on the heterogeneity of life in London by providing a comparative analysis of its 32 boroughs.

Data (1)

- Data used as variables in the clustering model come from London datastore, available publicly at: https://data.london.gov.uk/.
- London borough coordinates were extracted from the Wikipedia page available at: https://en.wikipedia.org/wiki/List_of_London_boroughs
- Data on venues in a chosen borough were extracted using Foursquare API.

Variables considered in the clustering model (1)

Variable	Description
green_space	Borough Green Space Surface (%) as of 2016
mean_income	Personal income by tax year in GBP as of 2018
median_income	Personal income by tax year in GBP as of 2018
total_area	Total Area (Hectares)
dwellings_p_hectare	Dwellings per hectare as of 2019
life_satisfaction	Mean score to a question "Overall, how satisfied are you with your life nowadays?", with a possible score from 0 to 10, as of 2018/2019
worhtwile	Mean score to a question "Overall, to what extent do you feel the things you do in your life are worthwile?", with a possible score from 0 to 10, as of 2018/2019

Variables considered in the clustering model (2)

Variable	Description
happiness	Mean score to a question "Overall, how happy did you feel yesterday?", with a possible score from 0 to 10, as of 2018/2019
anxiety	Mean score to a question "Overall, how anxious did you feel yesterday?", with a possible score from 0 to 10, as of 2018/2019
e_to_p	Ratio of House Prices to Earnings (full-time workers by place of work)
own_outright	Households owned outright (%) as of 2018
buying_w._mortg.	Households bought with mortgage or loan (%) as of 2018
rented	Households rented from local authority (%) as of 2018
rented_from_private	Households rented from private landlord (%) as of 2018
average_rent	Mean gross monthly rent paid in GBP as of 2019 Q1 (for all categories of dwellings)

Methodology

- In order to segment boroughs based on similar characteristics a **k-means clustering method** was used.
- K-means clustering is easy to implement and allows to arbitrarily choose the number of centroids/clusters (k).
- The chosen value of k is 3, which was supported by choropleth maps analysis and furthermore was considered to be analogous to a segmentation into low, medium and high income boroughs (such categories are often used for countries classification – see e.g. IMF).

Results (1)

Members of clusters

Cluster 0	Cluster 1	Cluster 2
Barking and Dagenham	Camden	Hackney
Barnet	Westminster	Hammersmith and Fulham
Bexley	Kensington and Chelsea	Haringey
Brent		Islington
Bromley		Lambeth
Croydon		Lewisham
Ealing		Newham
Enfield		Southwark
Greenwich		Tower Hamlets
Harrow		Wandsworth
Havering		
Hillingdon		
Hounslow		
Kingston upon Thames		
Merton		
Redbridge		
Richmond upon Thames		
Sutton		
Waltham Forest		

Discussion (1)

Based on the mean characteristics of clusters, Cluster 0 boroughs seem to be a good choice for first-time home owners. This is indicated by lower earnings to price ratio, lower income ratio and a high rate of properties bought with a mortgage (**buying_w._mortg.**). The last factor may hint to the fact that the property demand driven by investment motives is lower in these borrows, therefore it does not generate additional price pressures.

Discussion (2)

There are several factors that point to Cluster 1 boroughs as being the most attractive for property investment (i.e. buying residential property with the aim of gaining profit from renting it and not using it as the primary residence). Earnings to price ratio (**e_to_p**) is much higher than in the rest of the city, indicating a higher propensity and ability to invest spare income. A ratio of residential properties owned (**own_outright**) is similar to Cluster 0, but the share of those bought with a mortgage is much lower, supporting the hypothesis of higher ability (spare disposable income) and propensity to invest.

Discussion (3)

Cluster 2 boroughs are characterised by the lowest share of properties owned (**own_outright).** These are not as far away from the centre of the city as Cluster 0 boroughs, therefore the commuting times are shorter. This is a good alternative for young and mobile workers, who have no need to settle down yet. What is more, the **average_rent** in these cluster is not much higher than in Cluster 0 (at least in comparison to the centre of London).

Conclusion

- London boroughs can be characterised by a significant degree of heterogeneity.
- Cluster 0 boroughs are characterised by relatively low rent and higher share of green space.
- Cluster 1 features point to its attractiveness in terms of property investment.
- Young and mobile workers should benefit from medium rent prices and location of Cluster 2 boroughs.
- In terms of well-being of its residents, London boroughs were found to be very homogenous.