The Battle of Neighbourhoods in London

APPLIED DATA SCIENCE CAPSTONE – PART OF THE IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

Introduction

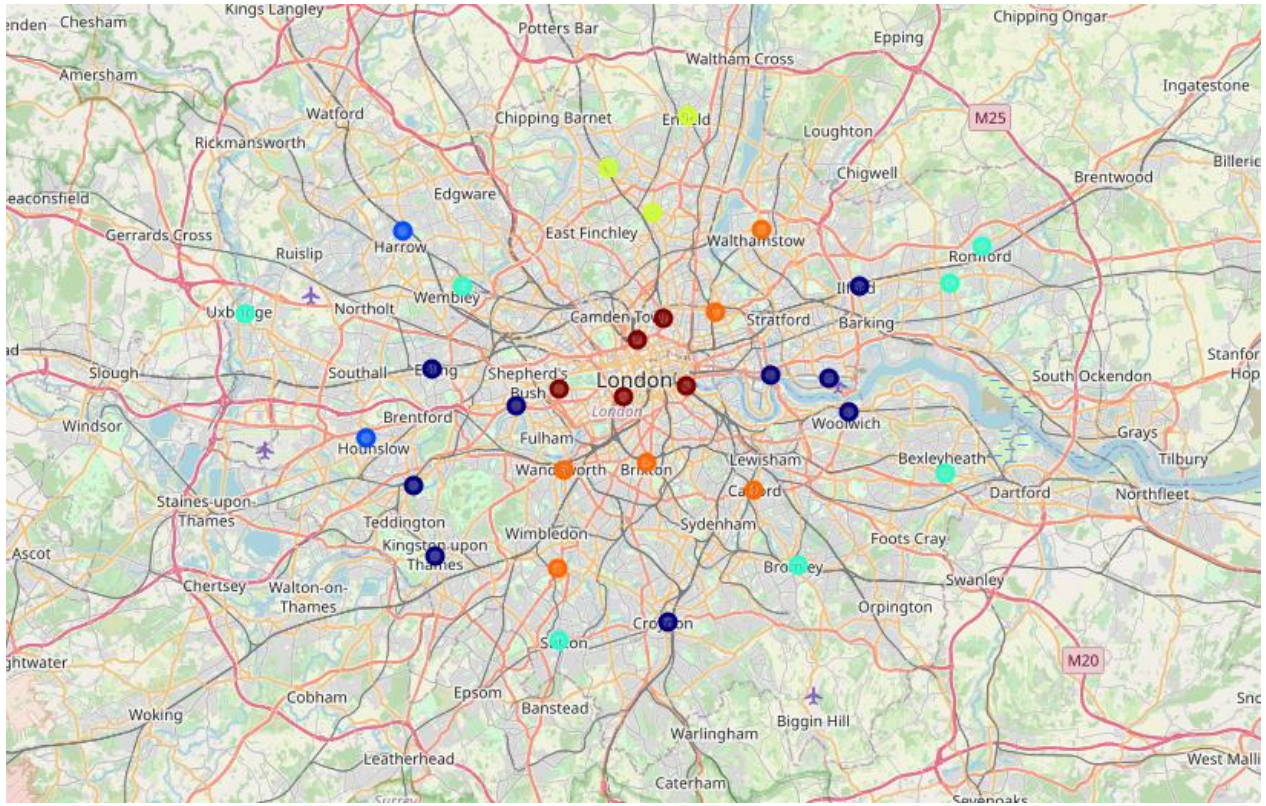
- London is one of the most diverse and populated city in the world with approximately 9 million people living in Greater London. London itself is divided into 32 boroughs plus the City of London.
- The purpose of this study is to analyse whether different clusters of boroughs, as determined by a k-means algorithm on Foursquare venue category data, also have similar social, demographic and economic attributes. For example, do boroughs of the same cluster have similar household income, well-being levels and crime rates? Further, how would London's boroughs be clustered based on socio-economic data?

Data Description and Cleaning

- Data for this study was obtained from Foursquare (venue data), Wikipedia (borough location data) and the London Datastore (socio-economic / borough geojson data).
- Data was readily available for all boroughs and no data cleaning was required.
- The City of London was excluded given it is not an official borough.
- Socio-economic data includes, for example, house price, income, crime rate and life satisfaction.

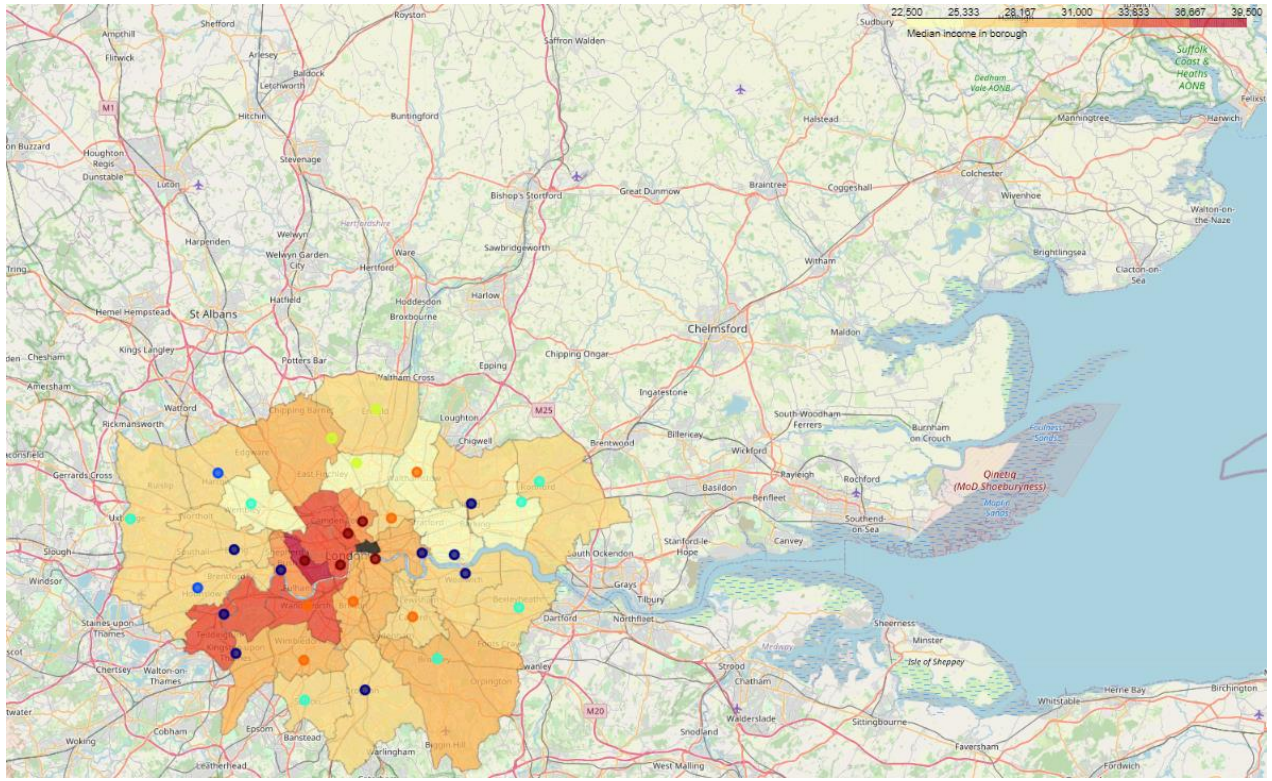
Clusters based on Foursquare Venues

- **Cluster 0 (red):** concentrated at the centre of London and includes, for example, Westminster and Camden.
- **Cluster 1 (purple):** divided into West London and East London boroughs with cluster 0 in-between.
- **Cluster 2 (blue):** concentrated in West London.
- **Cluster 3 (pink):** various (outer) boroughs around London.
- **Cluster 4 (yellow):** concentrated in North London.
- **Cluster 5 (orange):** concentrated in South-East London



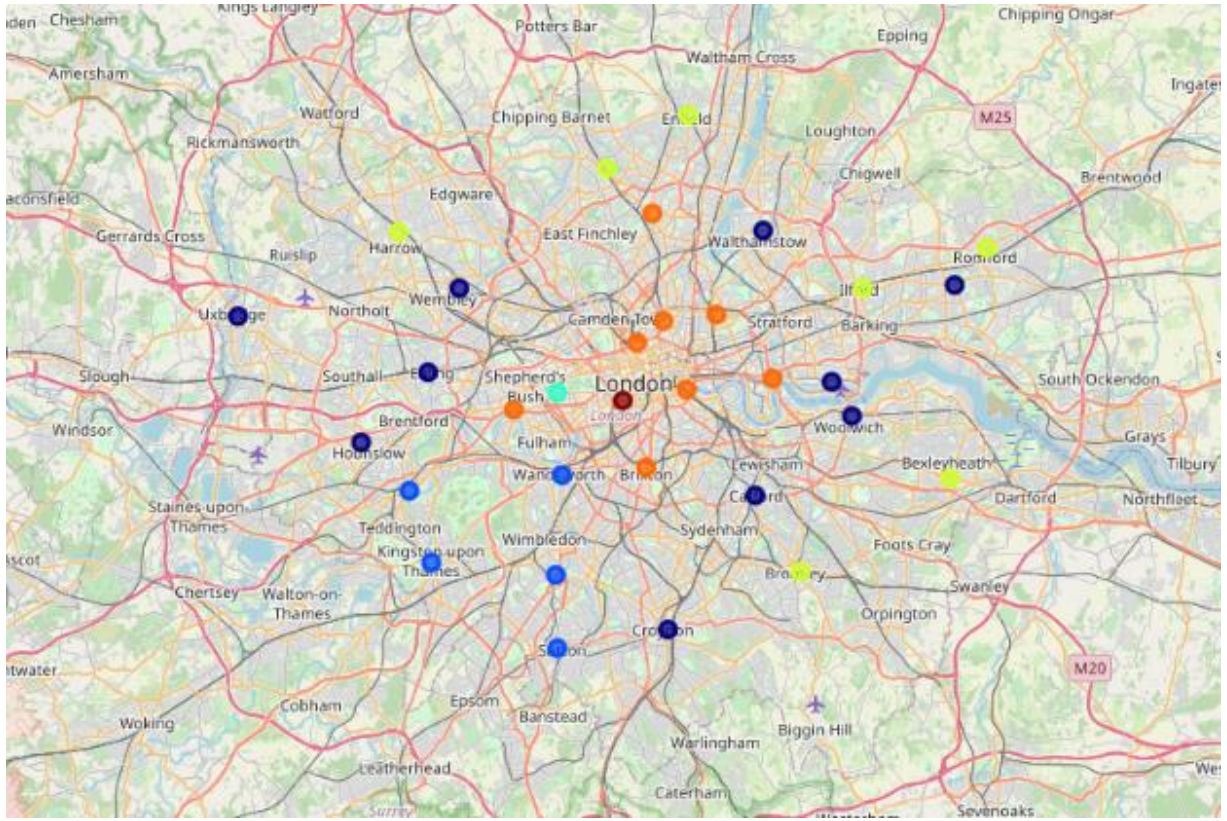
Clusters based on Foursquare Venues with Median Income Overlay

- There does not appear to be a clear relationship between median income and clusters.
- The same can be shown for median house price and crime rates (see report for details).



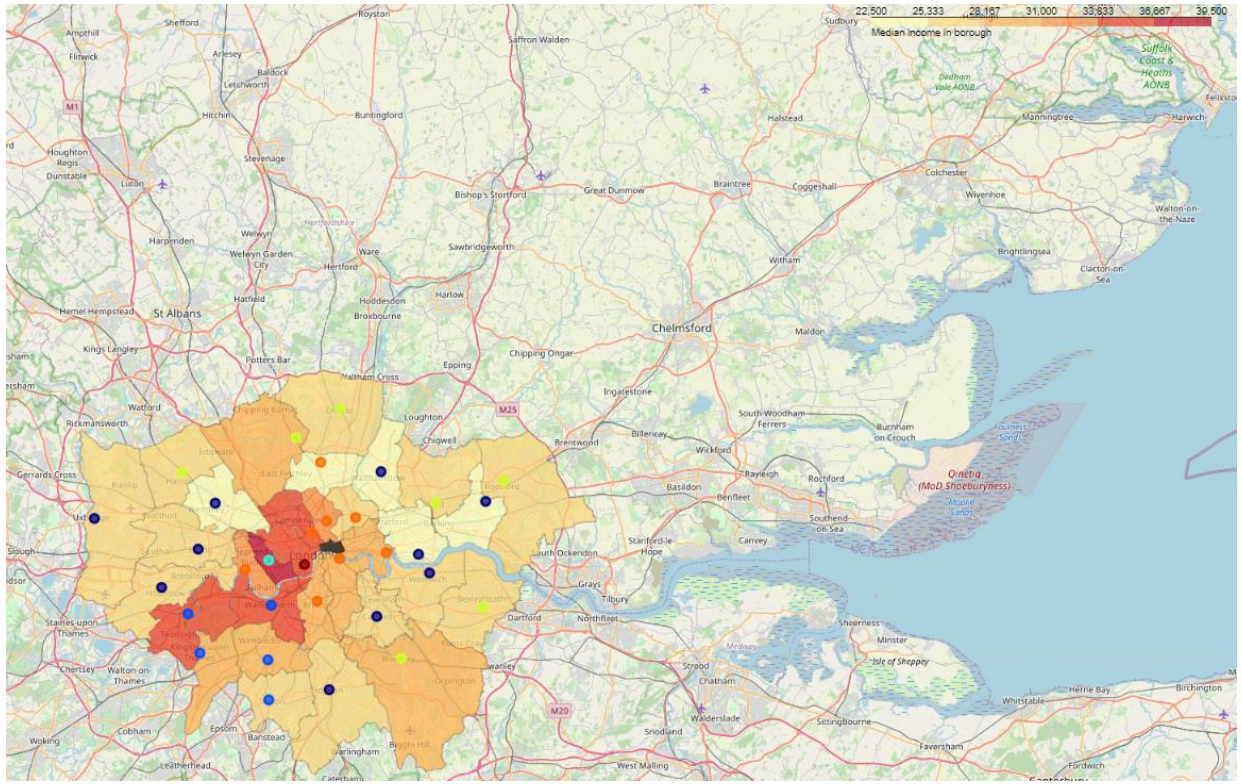
Clusters based on socio-economic data

- **Cluster 0 (red):** single borough, only consists of Westminster.
- **Cluster 1 (purple):** some of the outer boroughs.
- **Cluster 2 (blue):** concentrated in (South-)West London.
- **Cluster 3 (pink):** single borough, only consists of Kensington and Chelsea.
- **Cluster 4 (yellow):** outermost boroughs in (North-/South-)East London.
- **Cluster 5 (orange):** mainly consists of central London boroughs



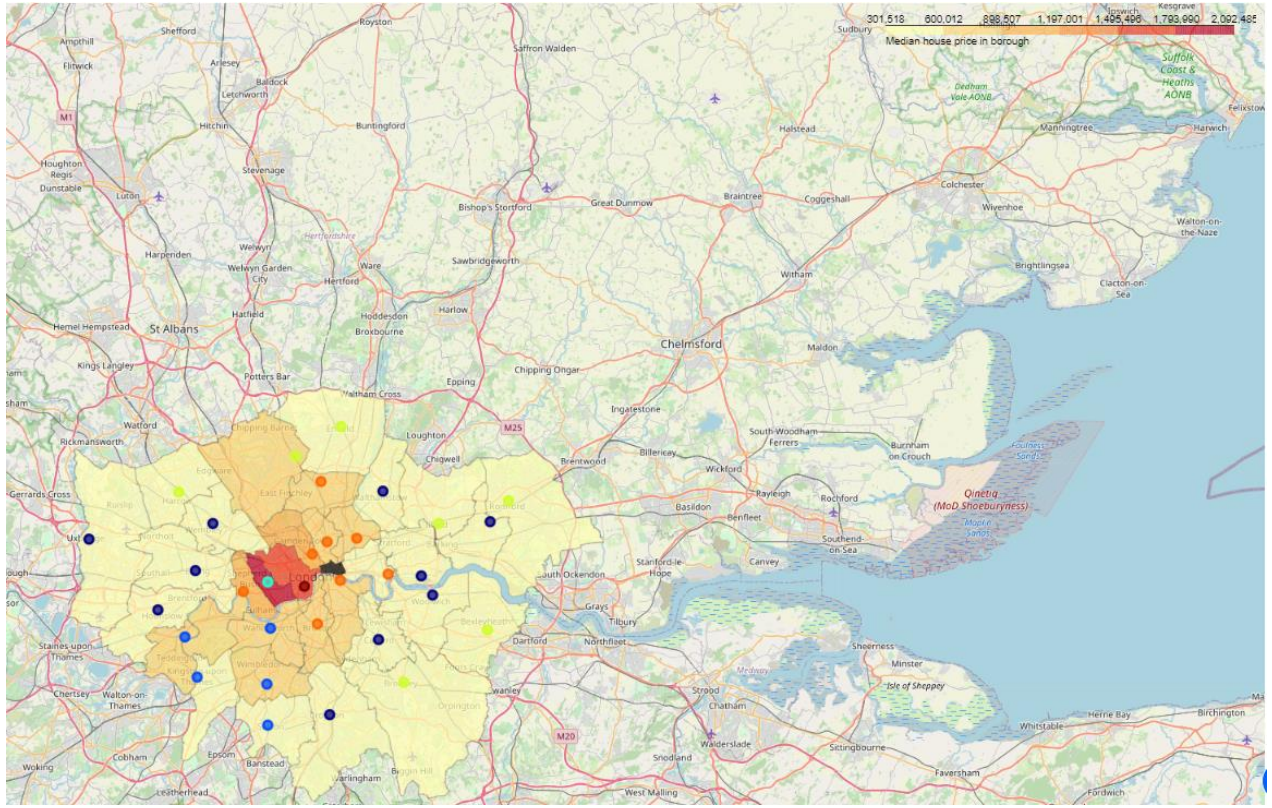
Clusters based on socio-economic data with Median Income Overlay

- There does not appear to be a clear trend for different clusters.
- In general, central London boroughs and a handful of West London boroughs have the highest median income.
- Median income is lower in South and East London.



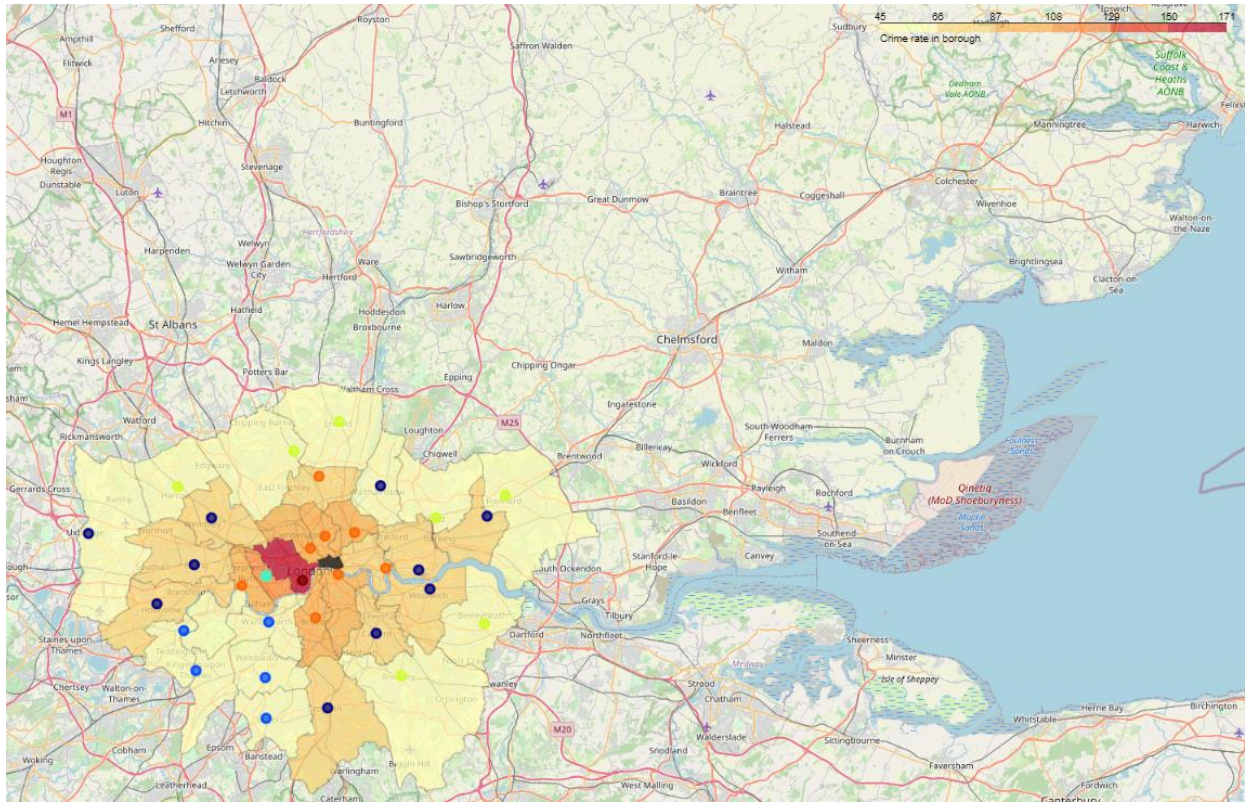
Clusters based on socio-economic data with Median House Price Overlay

- Westminster (red) and Kensington and Chelsea (pink) stand out with the highest median house prices.
- Blue and orange boroughs have (mostly) similar levels of median house prices.
- Yellow and purple boroughs have (mostly) the lowest median house prices.



Clusters based on socio-economic data with Crime Rate Overlay

- Boroughs within in each cluster appear to have similar crime rates.
- In most cases, purple boroughs appear to have higher crime rates than blue boroughs
- Westminster (red) and Kensington and Chelsea (pink) stand out with significantly higher crime rates.



Conclusion

- Clustering London's boroughs based on Foursquare venue data does not necessarily reflect the underlying socio-economic attributes.
- Given that venue data is dynamic, clusters change regularly thereby making it difficult to reach definitive conclusions.
- Future studies should look at additional socio-economic attributes as well as combining these attributes with, for example, Foursquare venue data.