DATA-LED GUIDE TO HOME FROM HOME

By Shirley Ng – August 2020

Finding a suitable place to be ‘home from home’

Background: Hong Kong is going through a turbulent times. People in this vibrant metropolitan city fear the loss of freedoms. Following the Chinese government’s decision to impose a new National Security Law, the UK government has committed to open a new immigration route to British National (Overseas) passport holders in Hong Kong.

Problem: Despite the fact that many Hong Kongers are well-travelled for holidays, it is still a challenge to understand various aspects of livelihood in a new country quickly for an uprooting move. There are numerous videos, social media posts, blogs and forums flowing with varying subjective opinions and information of mixed quality.

Interest: Selecting a suitable location to settle down in a 'home from home' is an important and complex decision. There is a need to analyse more recent and widely trusted data sources to provide an objective data-led view to help them make an informed decision. This project will cover these criteria.

- Common Venues
- Property Price
- Crime Rate
- Jobs
- Education

Data acquisition and cleaning

Data acquisition: This project analyses 42 cities in England that have data across all criteria. Data from Four Square, government sources like Office of National Statistics (ONS) and Department of Education as well as some websites are used.

Data processing and cleaning: Data are downloaded from different sources as listed below and scraped from webpages. Some of the data are processed by Microsoft Excel VLOOKUP function to extract data of relevant cities. Since several data sources are based on local authority as unit instead of city name, data cleaning involves amending some of the naming like ‘Kingston upon Hull’ to ‘Hull’. The list started with 51 cities but eventually trimmed down to 42 cities to only keep those that have complete data across all criteria.

Common Venues
Foursquare
August 2020

Property Price
HM Land Registry
January 2020

Crime Rate
ONS
March 2020

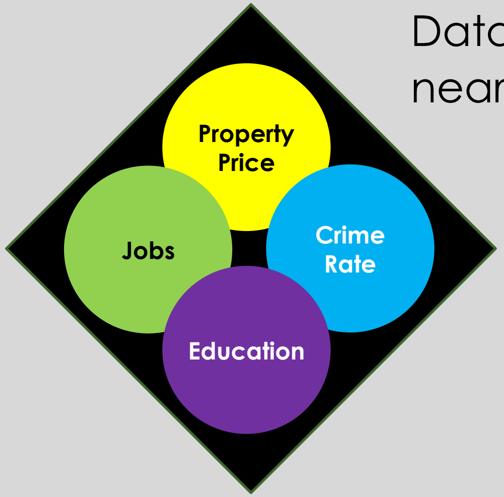
Jobs
Nomis
December 2019

Education
ONS
2018-2019

Data analysis

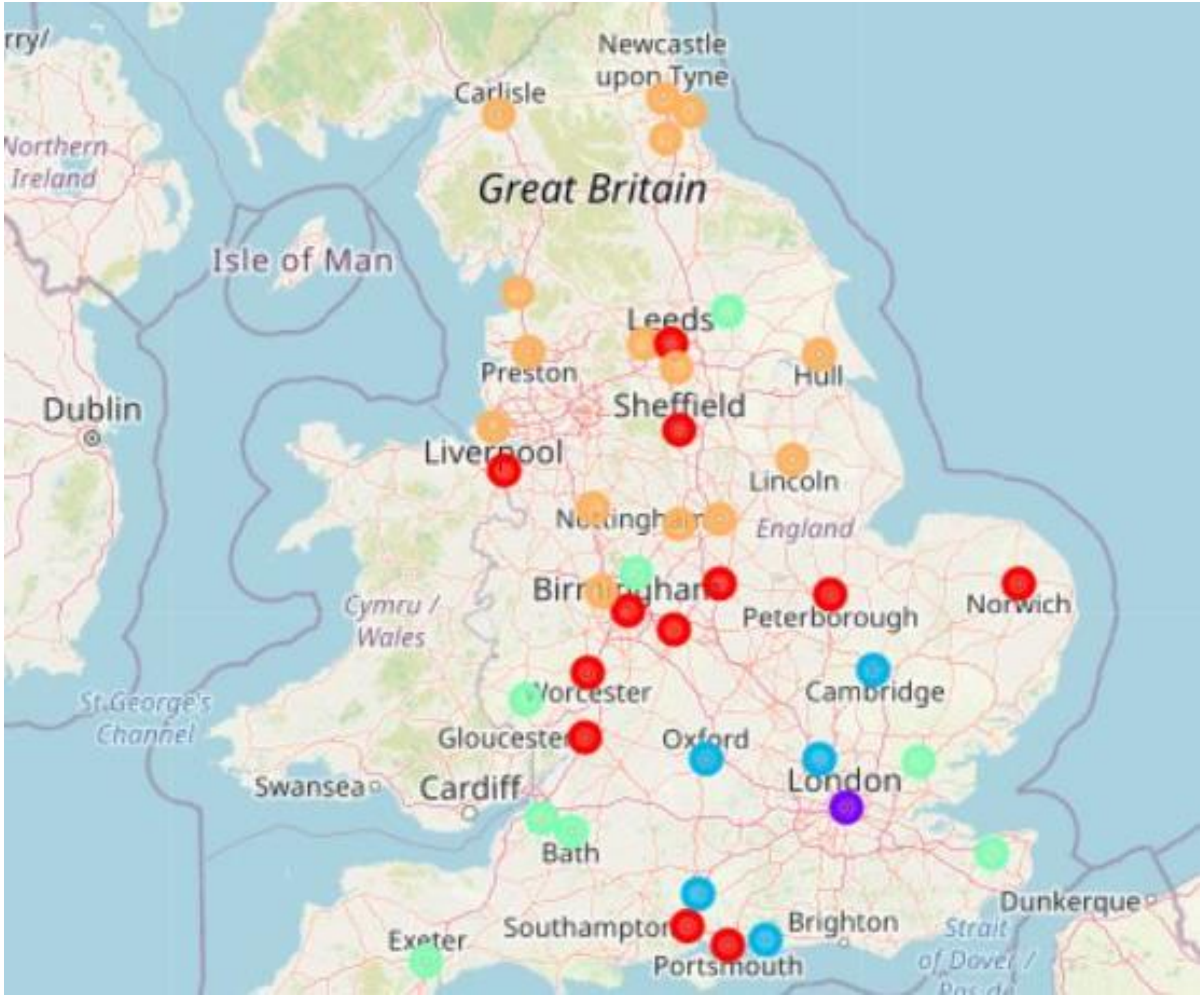
Data of up to 100 venues within 500m radius of each city are extracted from **Foursquare**. **One Hot Encoding** is used to transform nearby venue category into binary data.

	City	Accessories Store	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Asian Restaurant	Auto Garage	BBQ Joint	...	University	Vegetarian / Vegan Restaurant	Video Game Store	Video Store	Vietnamese Restaurant
0	Bath	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
1	Bath	0	0	0	0	0	0	0	0	0	...	0	1	0	0	0
2	Bath	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	Bath	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0



Data from **government sources** are combined into a data frame that is merged with the nearby venue data for a **k-means cluster segmentation**.

	City	Total crime	Violent crime	Theft offences	Property_Jan_2020	Employment rate	(MeetExp, mean)	(ATT8SCR, mean)
0	Bath	65.1	20.5	23.5	330975.0	0.731638	0.692414	39.353846
1	Birmingham	100.5	35.7	36.2	189161.0	0.640017	0.635331	45.800971
2	Bradford	136.1	53.2	38.8	133763.0	0.627727	0.617545	41.223529
3	Brighton and Hove	100.5	31.4	37.8	368224.0	0.769443	0.840000	NaN
4	Bristol	114.1	34.7	40.3	285296.0	0.778697	0.658955	43.990566



42 English cities into 5 clusters

Cluster 1: Birmingham, Chester, Coventry, Gloucester, Leeds, Leicester, Norwich, Peterborough, Plymouth, Portsmouth, Sheffield, Southampton, Worcester.

Cluster 2: London.

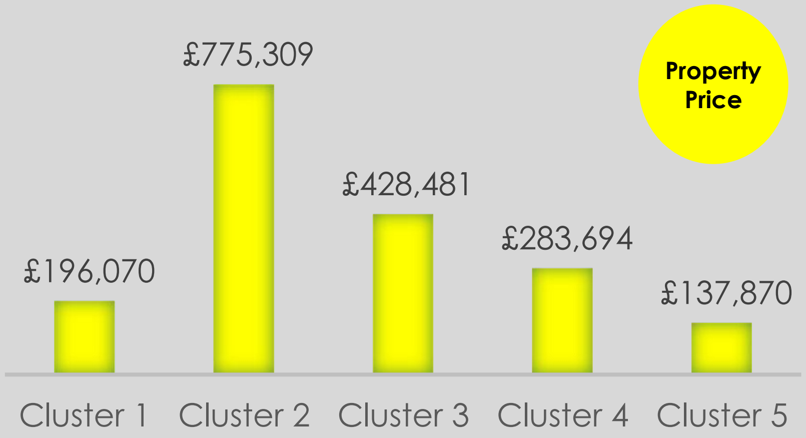
Cluster 3: Cambridge, Chichester, Oxford, St Albans, Winchester.

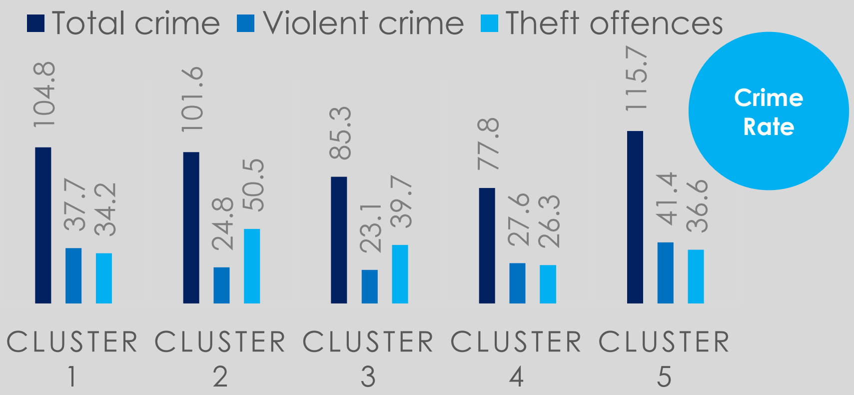
Cluster 4: Bath, Bristol, Canterbury, Chelmsford, Exeter, Hereford, Lichfield, York.

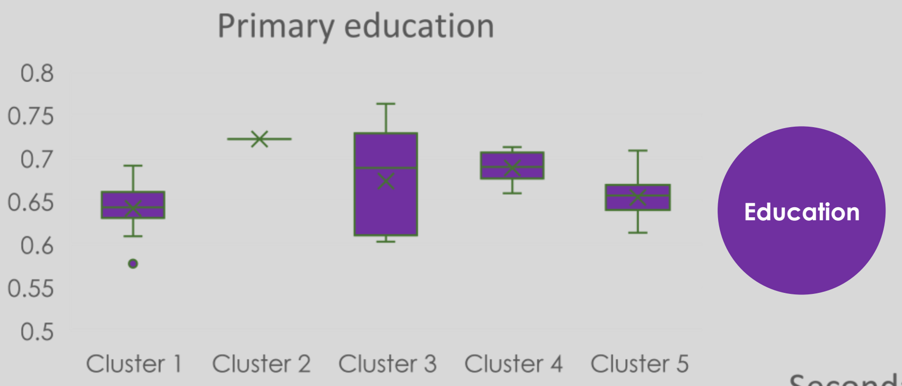
Cluster 5: Bradford, Carlisle, Derby, Durham, Hull, Lancaster, Lincoln, Liverpool, Newcastle upon Tyne, Nottingham, Preston, Stoke-on-Trent, Sunderland, Wakefield, Wolverhampton.

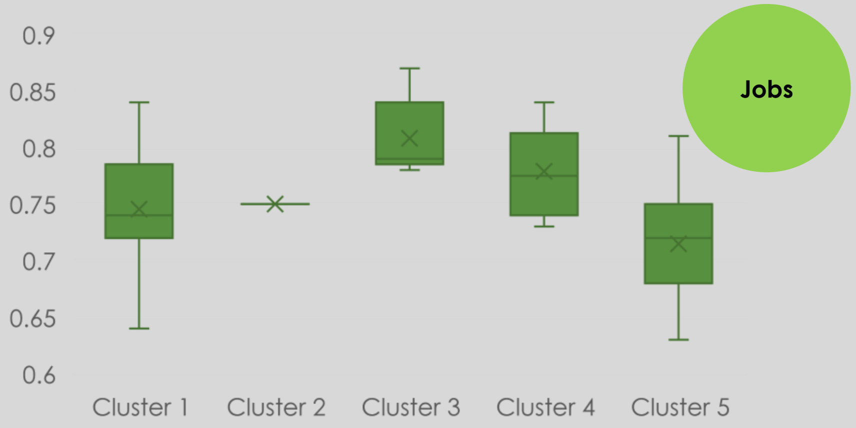
Examine the clusters

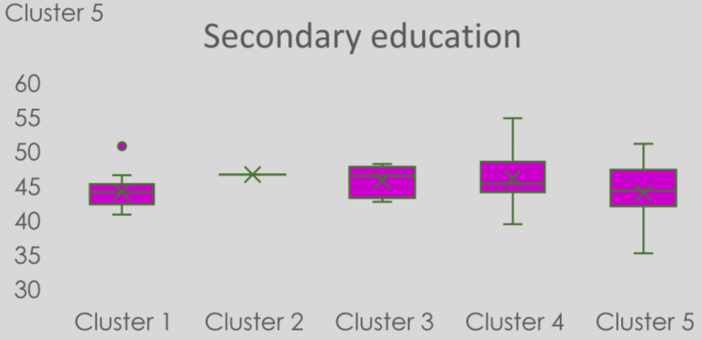
The mean value of each cluster's data are compared in charts.











Conclusion

Conclusion: The k-means cluster segmentation confirms that London is 'one of a kind'. Therefore, the correlation analysis is performed across all clusters, as well as excluding London to reveal any correlations hidden by the unique London data.

As shown in the table below, high violent crime is a strong predictor for low property price, low employment rate and lower secondary school performance. In contrast, employment rate and property price index are positively correlated outside of London. No surprise in the strong positive correlation between primary and secondary school performance as well.

	Total crime	Violent crime	Theft offences	Property price	Employment rate	Primary Education	Secondary Education
Property							
All	-18%	-76%	80%	-	28%	89%	84%
Excl. LDN	-79%	-96%	20%	-	98%	61%	79%
Employment							
All	-89%	-82%	-12%	28%	-	25%	56%
Excl. LDN	-89%	-99%	1%	98%	-	66%	86%
Pri. Edu							
All	-35%	-75%	53%	89%	25%	-	94%
Excl. LDN	-86%	-78%	-46%	61%	66%	-	94%
Sec. Edu							
All	-64%	-91%	36%	84%	56%	94%	-
Excl. LDN	-98%	-92%	-39%	79%	86%	94%	-