

LOCATION, LOCATION, LOCATION

Finding the best borough in London

Carlos Noble Jesus

PROBLEM...

London is a huge city
comprised of 32 boroughs
(excluding City of London)

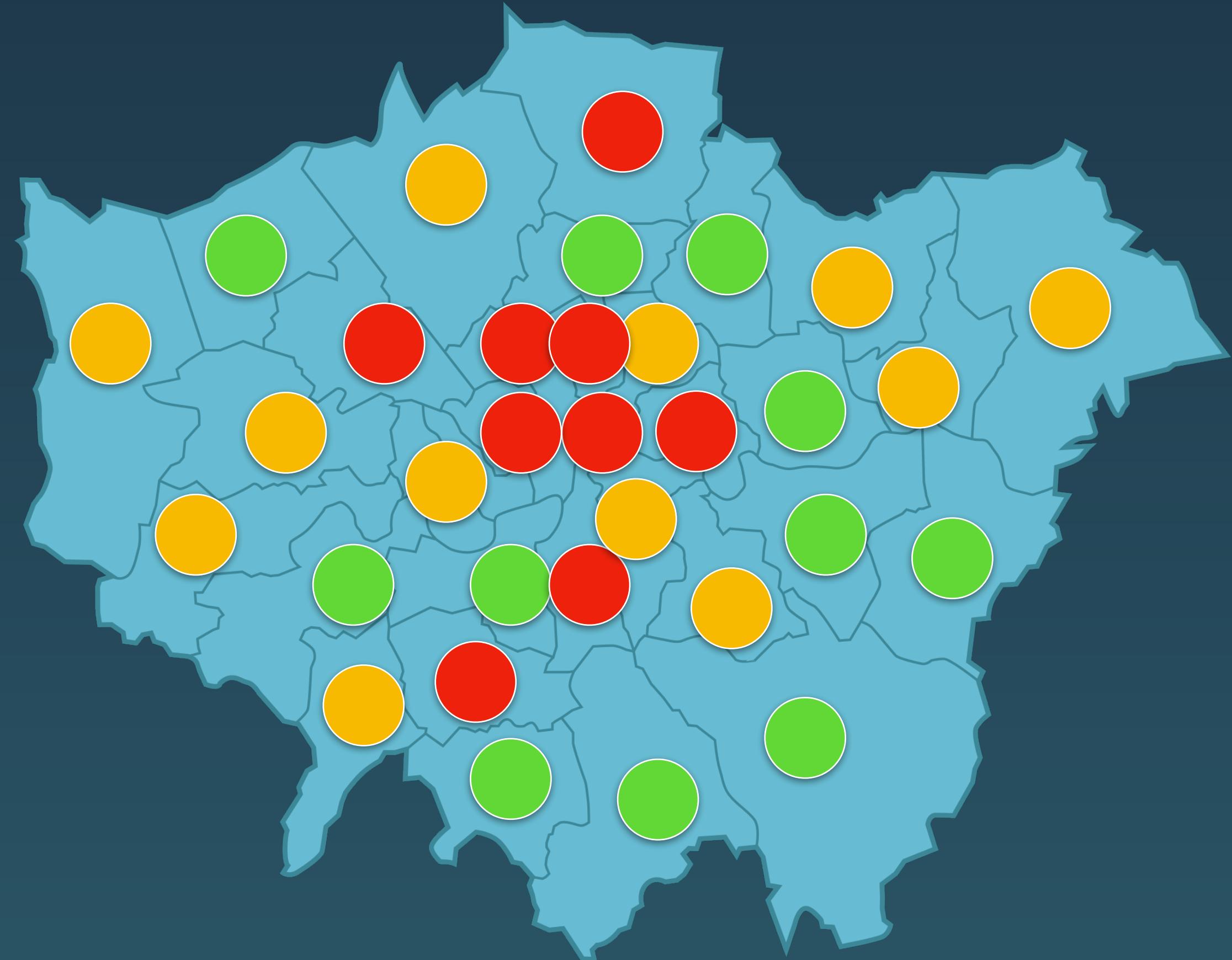
How can people moving to London
to live or carry out business
decide where to locate?



SOLUTION!

Build a model that clusters the boroughs in London based on key features

The clusters can be used to determine the suitability of a borough based on the individual or company's needs



DATA ACQUISITION & CLEANING

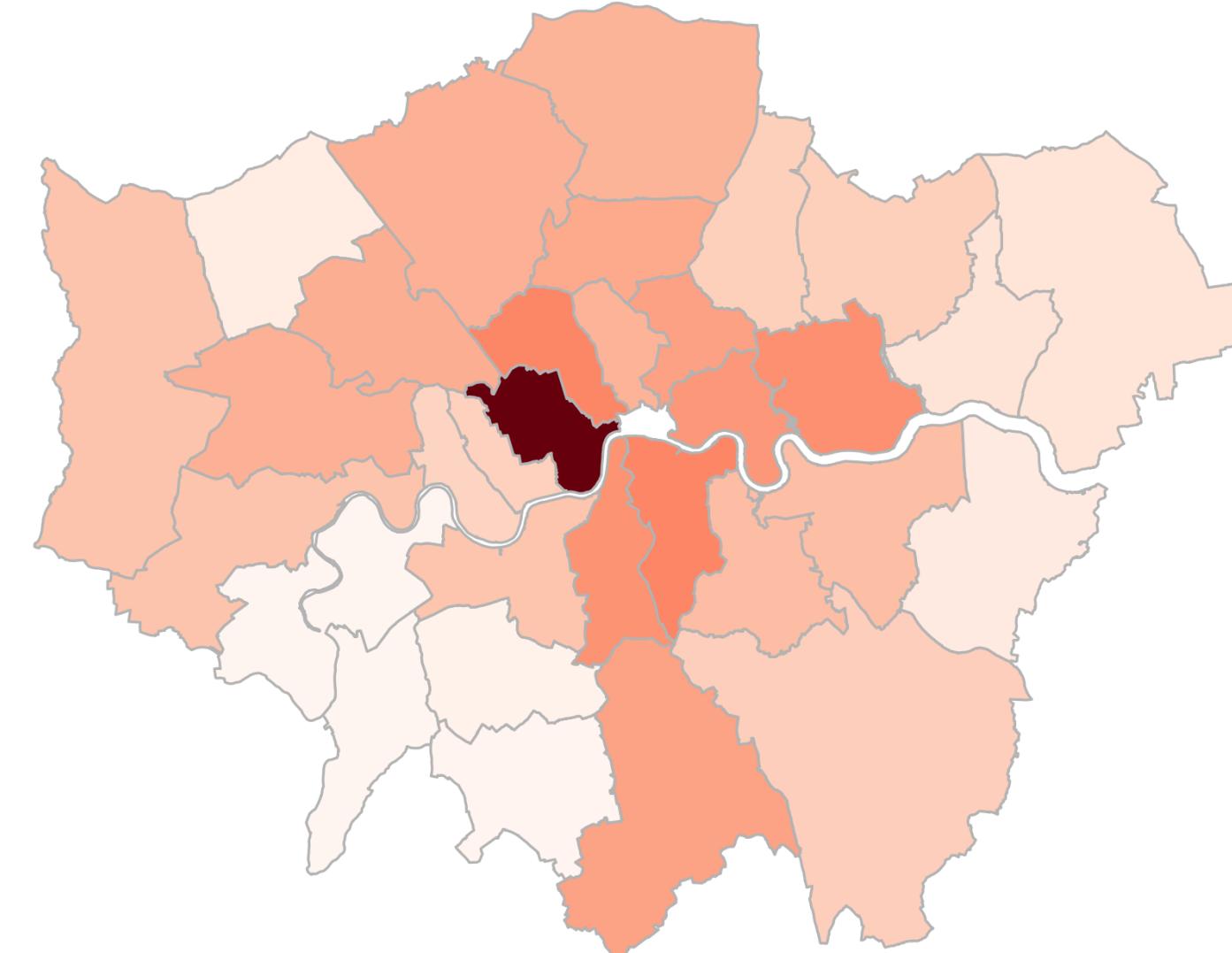
Data from several sources were acquired:

- Metropolitan Police
- London Datastore
- Wikipedia (parsing table)
- Foursquare API

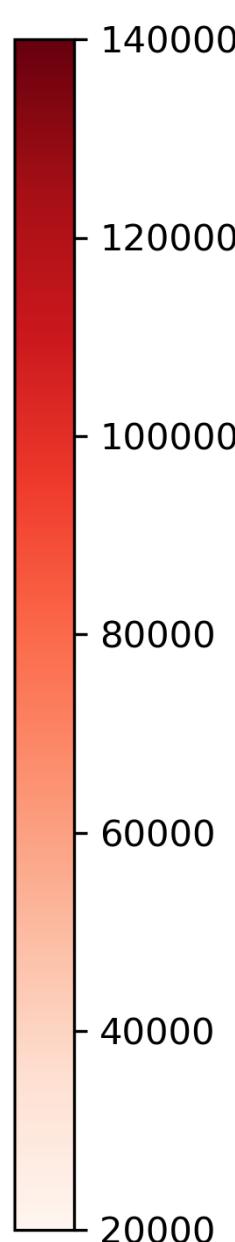
Data was cleaned and merged to produce one final dataframe

CRIME DATA

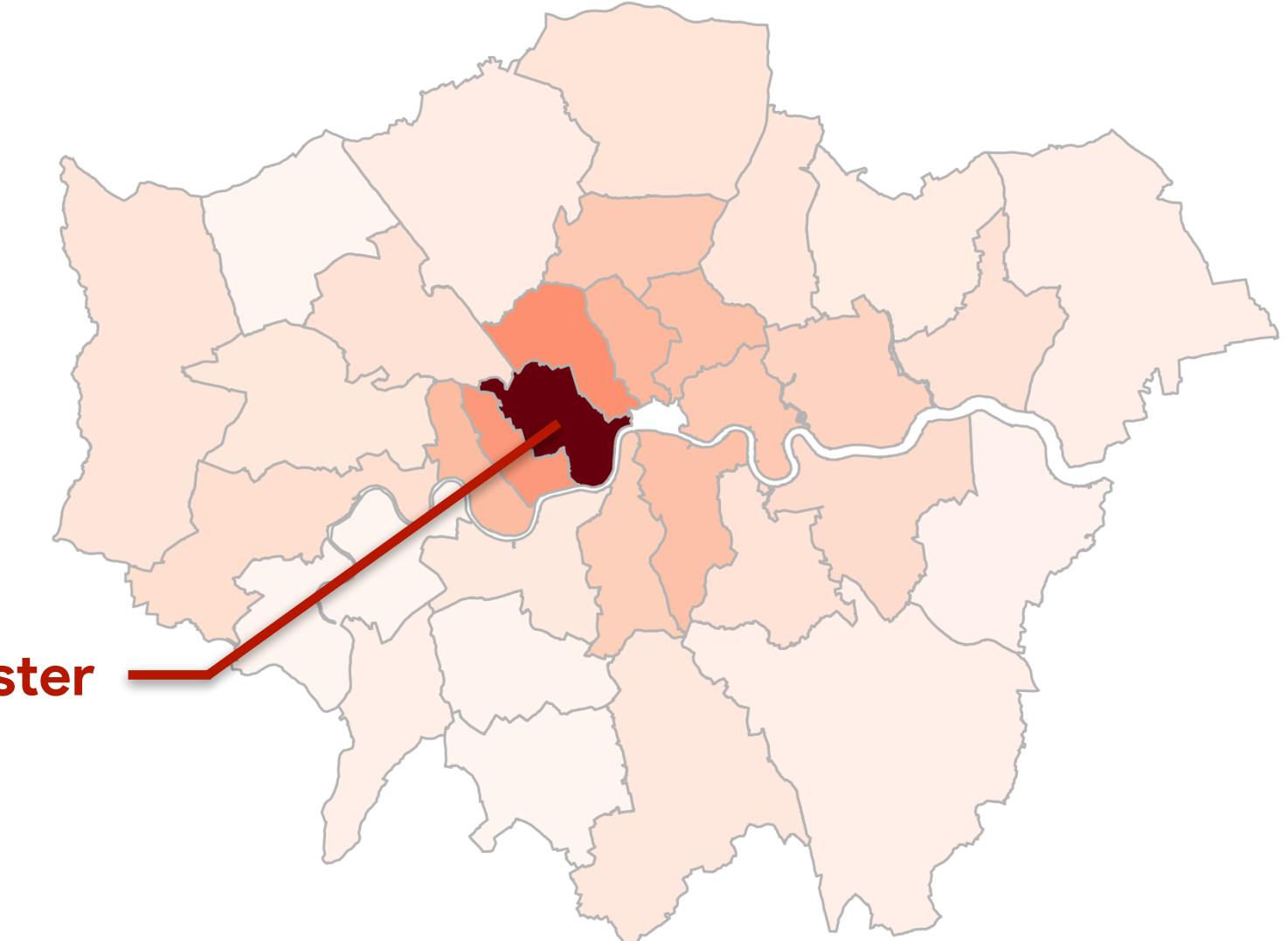
Total number of crimes



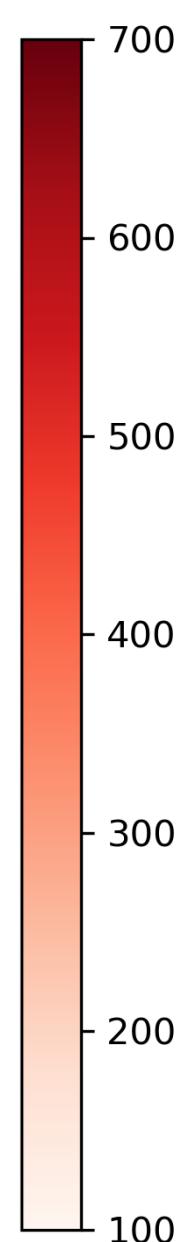
Source: Metropolitan Police, 2020



Average number of crimes per 1000 people



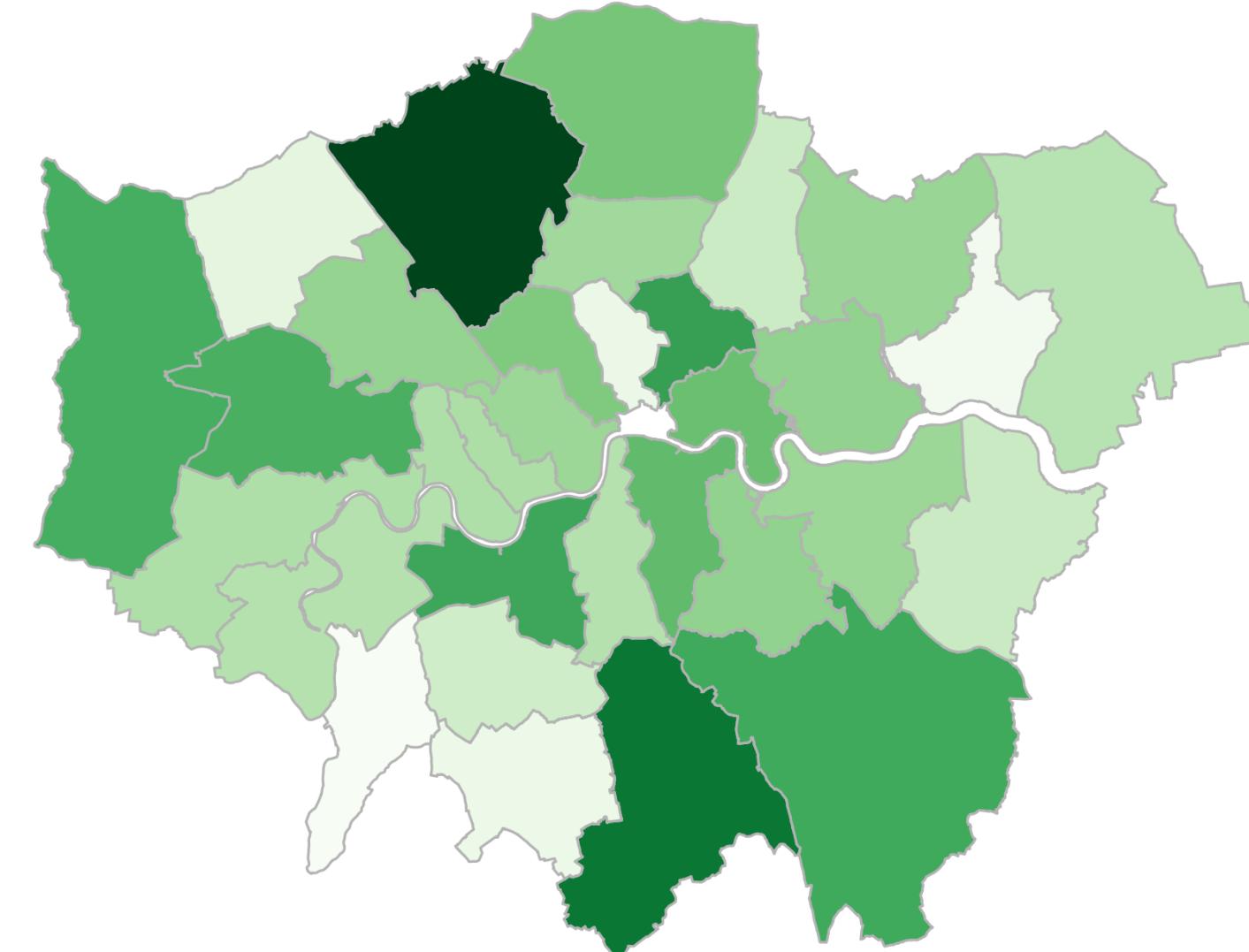
Source: Metropolitan Police, 2020



Normalising crime data by population shows that the central boroughs feature a higher crime rate than the outer boroughs

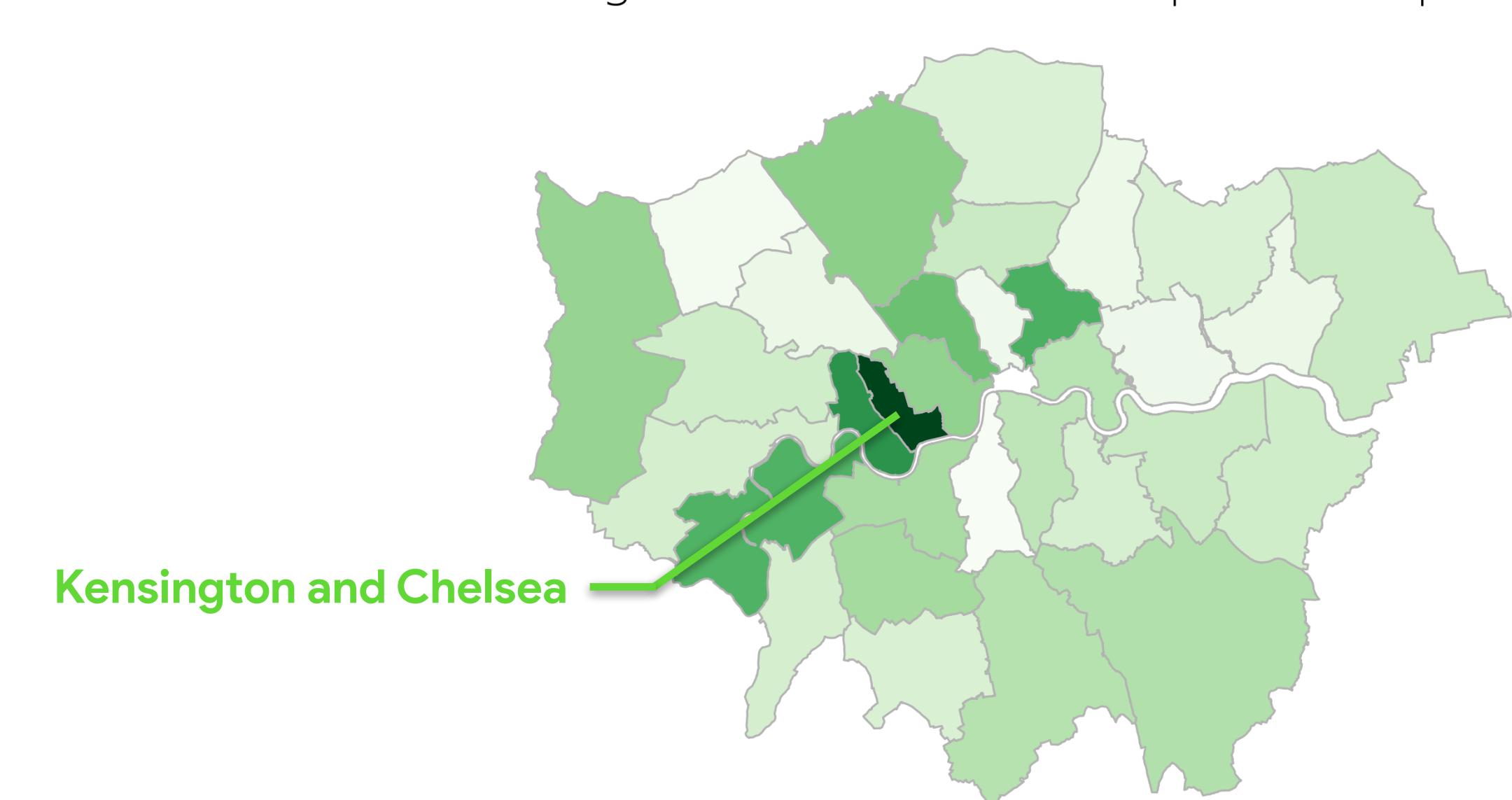
SCHOOL DATA

Total number of schools



Source: Greater London Authority, 2016

Average number of schools per 1000 people

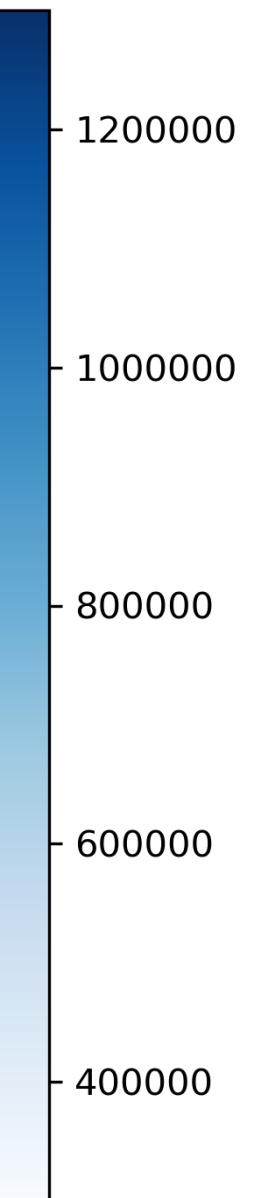
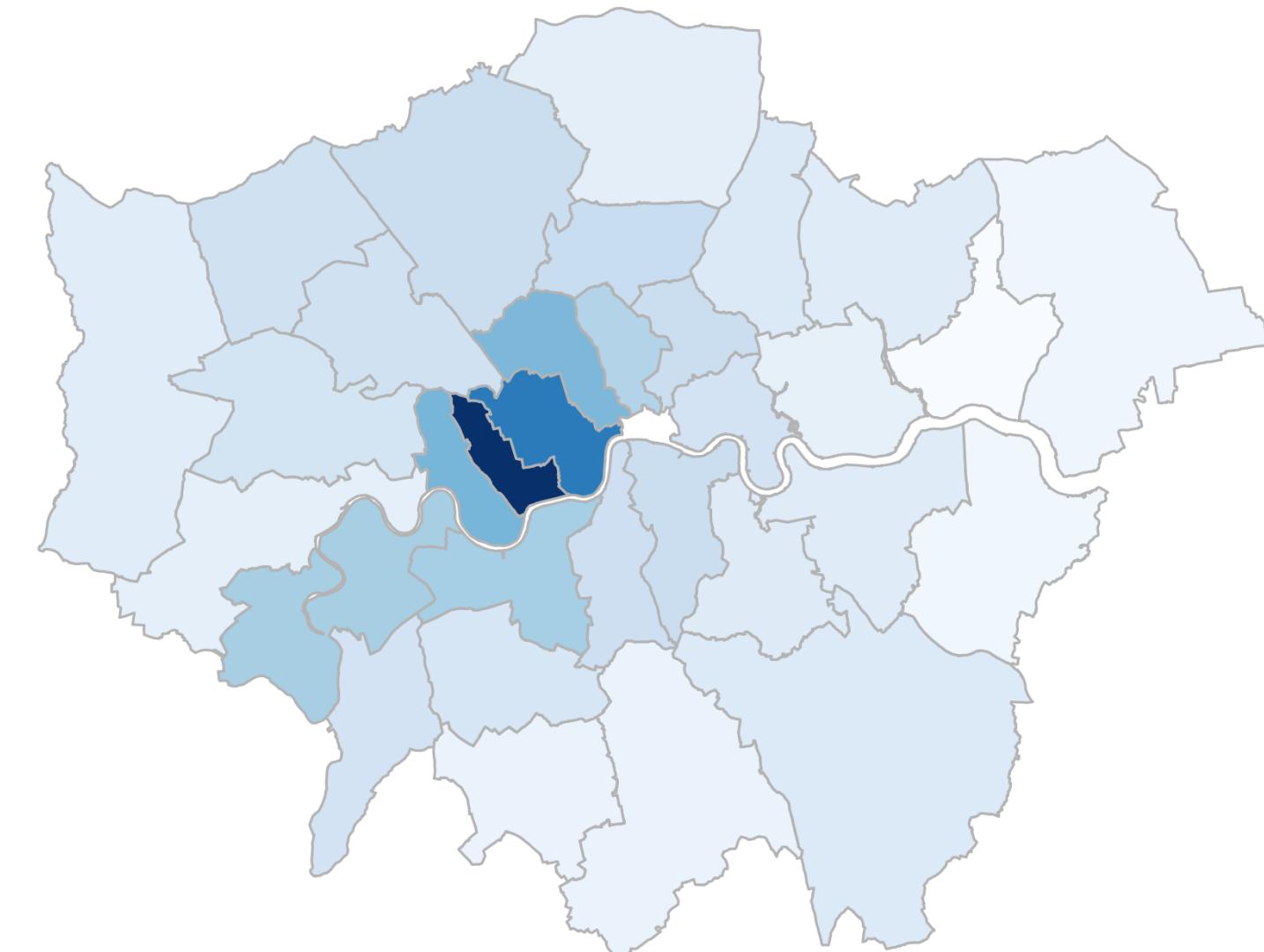


Source: Greater London Authority, 2016

Normalising the number of schools by population also illustrates that the total number of schools may not be a reliable metric

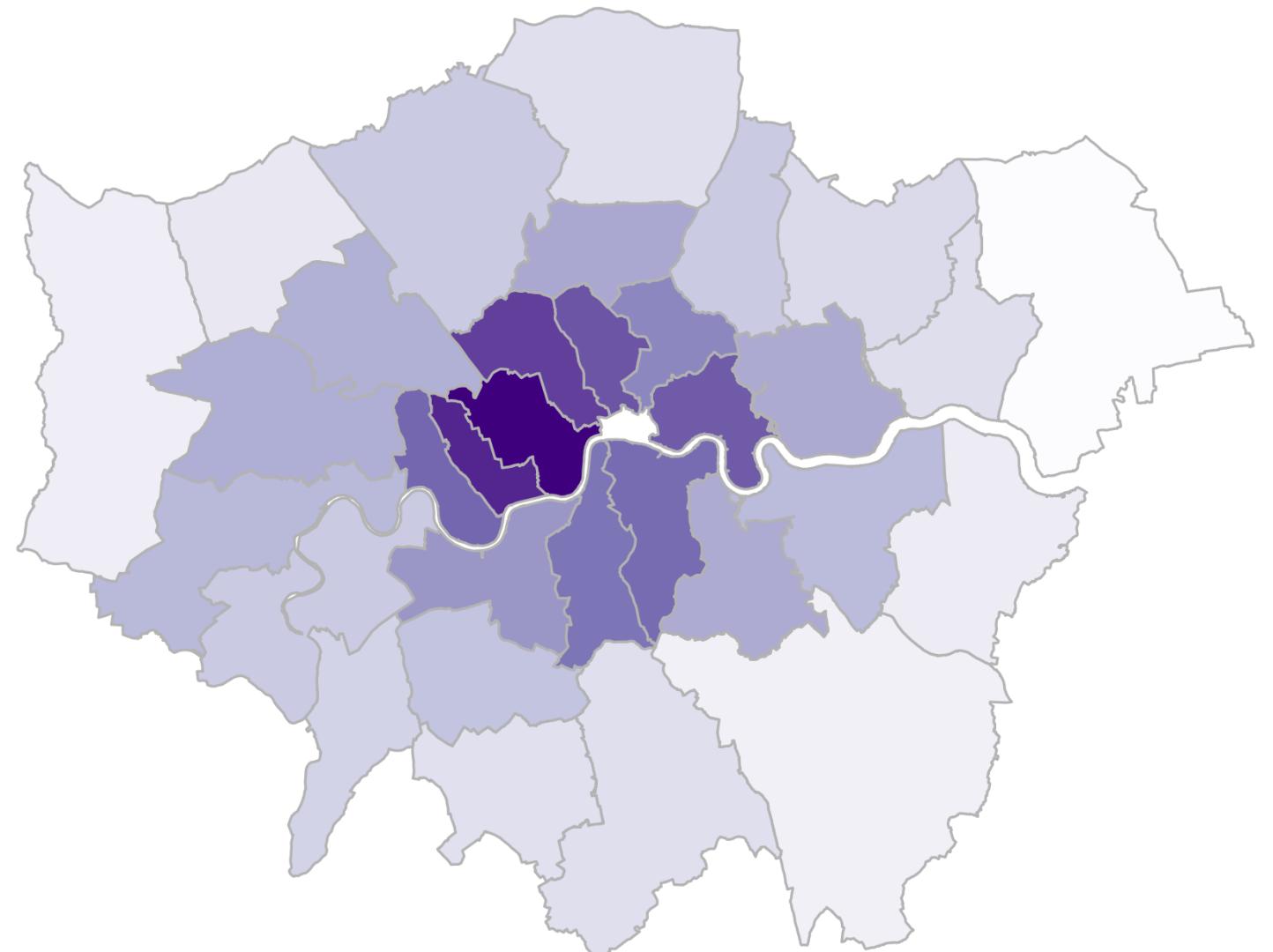
OTHER FEATURE DATA

Median House Price, £ (2017)



Source: London Datastore, 2018

NO₂ emissions, µg/m³



Source: London Datastore, 2018

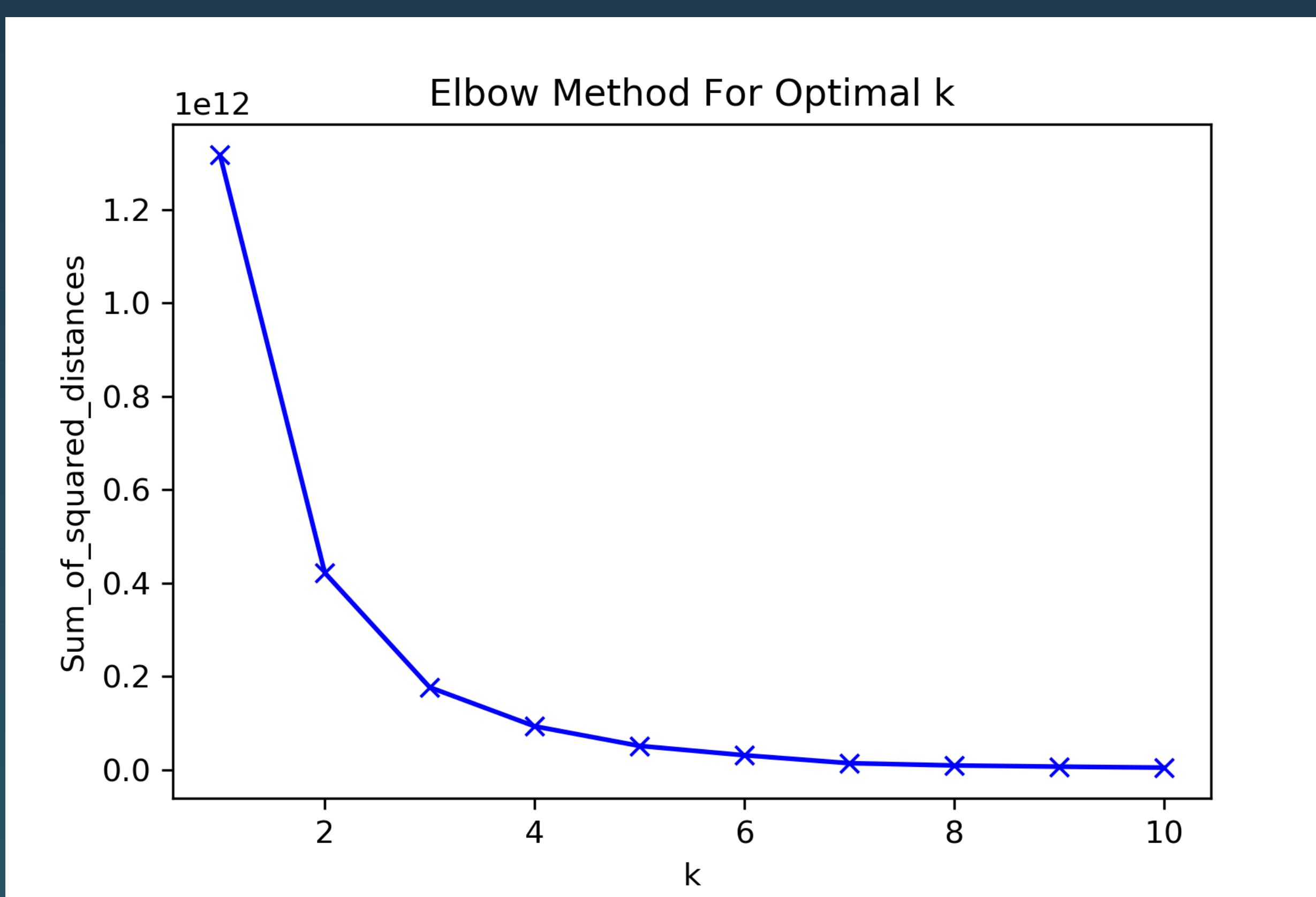
House prices and air pollution is also highest in the central boroughs

CLUSTERING THE BOROUGHS

K-Means clustering

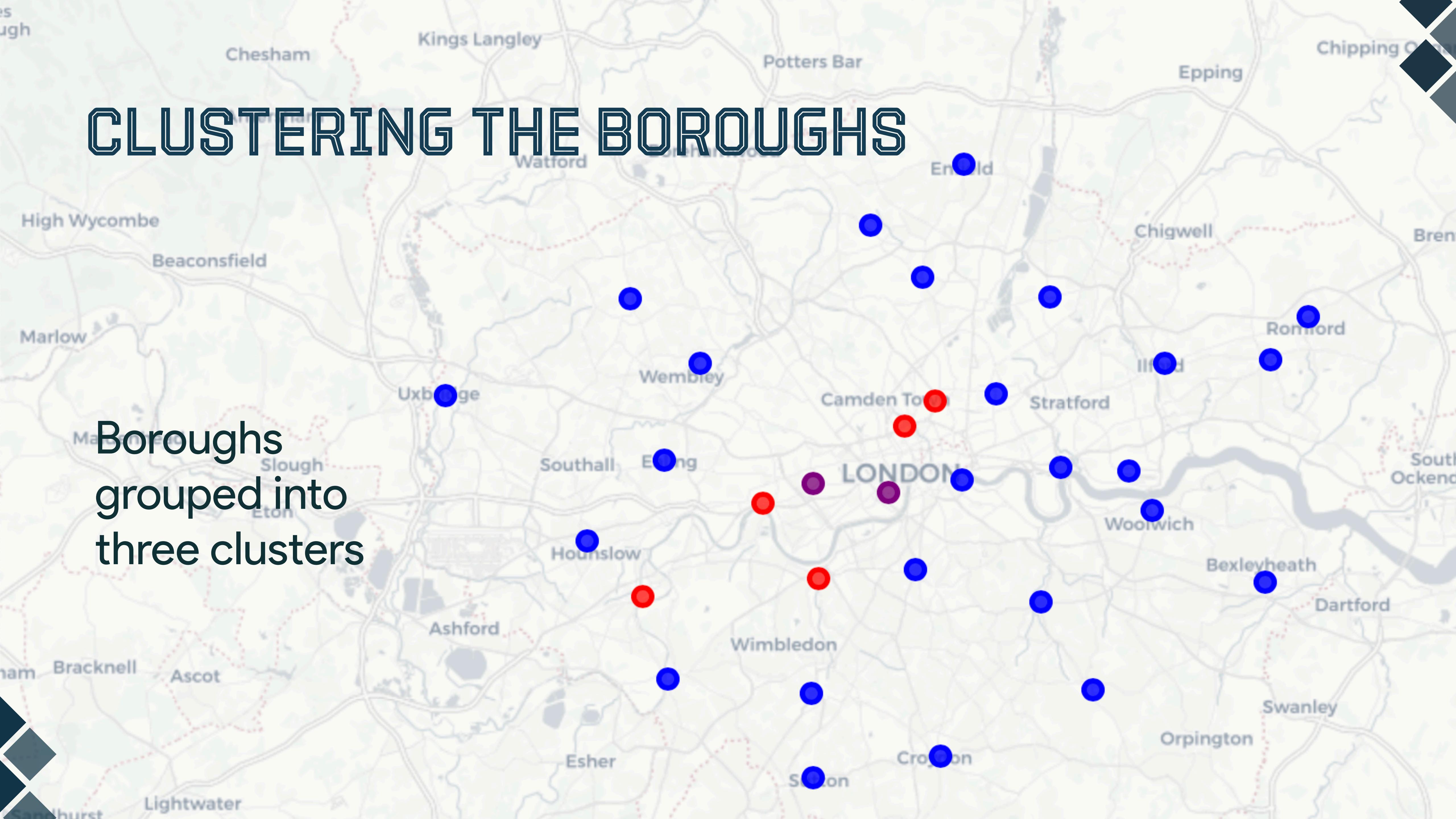
Subset of features used
for model

Elbow method to obtain
optimal k value



CLUSTERING THE BOROUGHS

Boroughs
grouped into
three clusters



ANALYSING THE CLUSTERS

The correlation between the cluster labels (0, 1 or 2) and the model features was calculated

Median House Price	0.610855
exceed_N02	0.487950
Schools per 1000	0.479676
Pop Density	0.382992
Supermarkets	0.365960
Crimes per 1000	0.341027
Hospitals	0.295030
Parks	0.278380

Median house price seems to be the main feature influencing the clustering

ANALYSING THE CLUSTERS

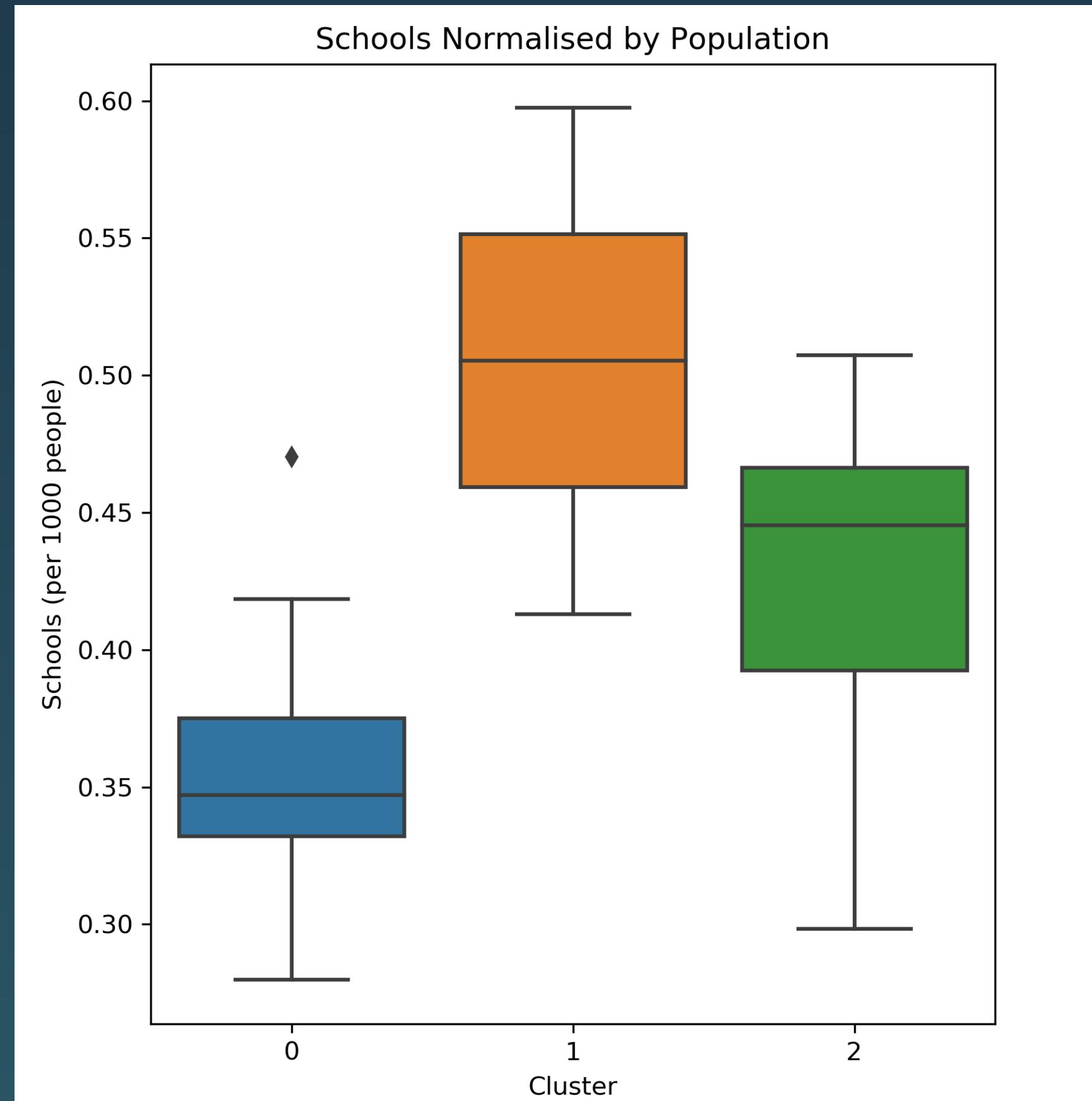
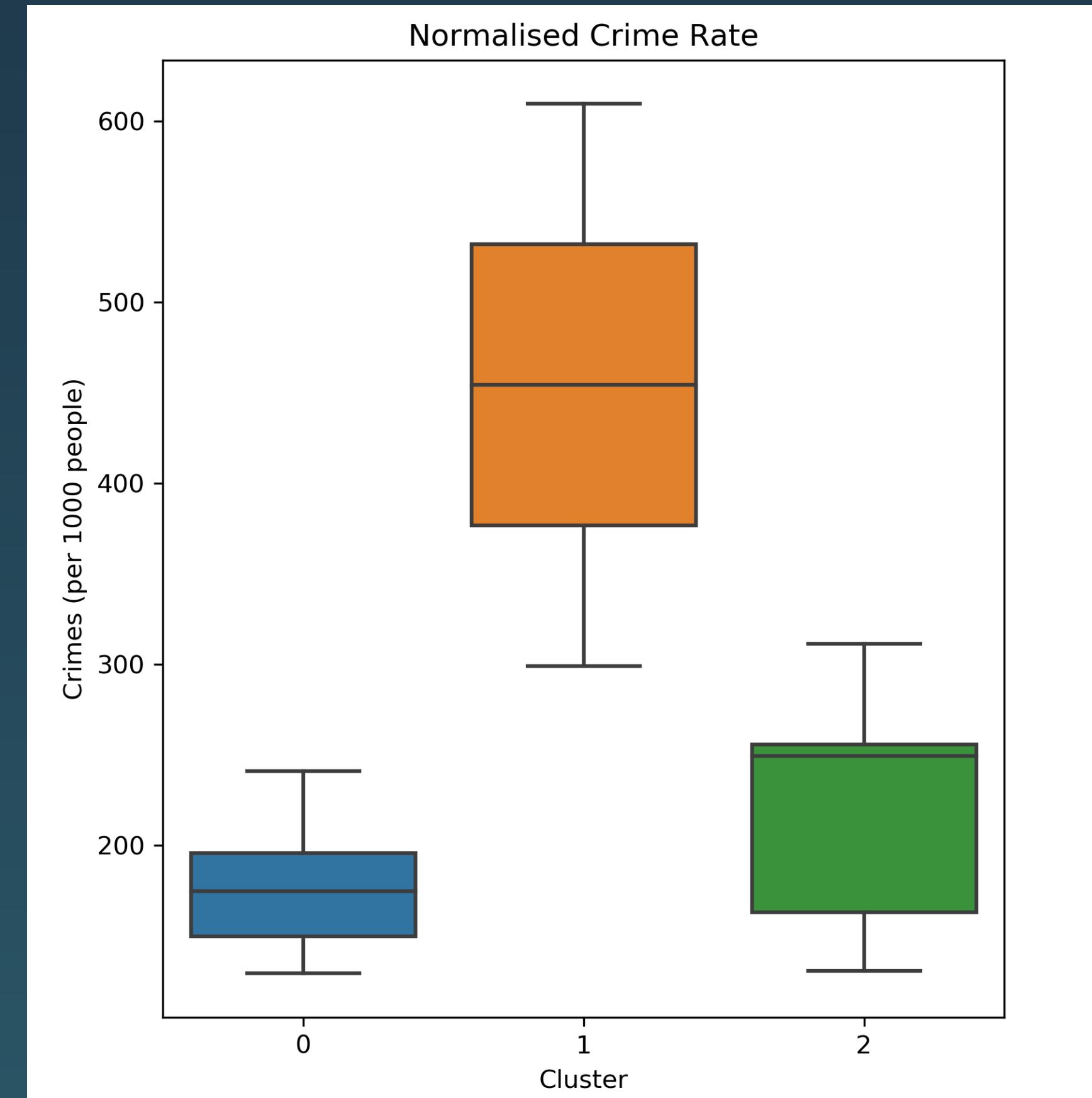
Box plots for some of the highest correlating features were made to visualise the differences between clusters



Median house price was significantly different between each cluster

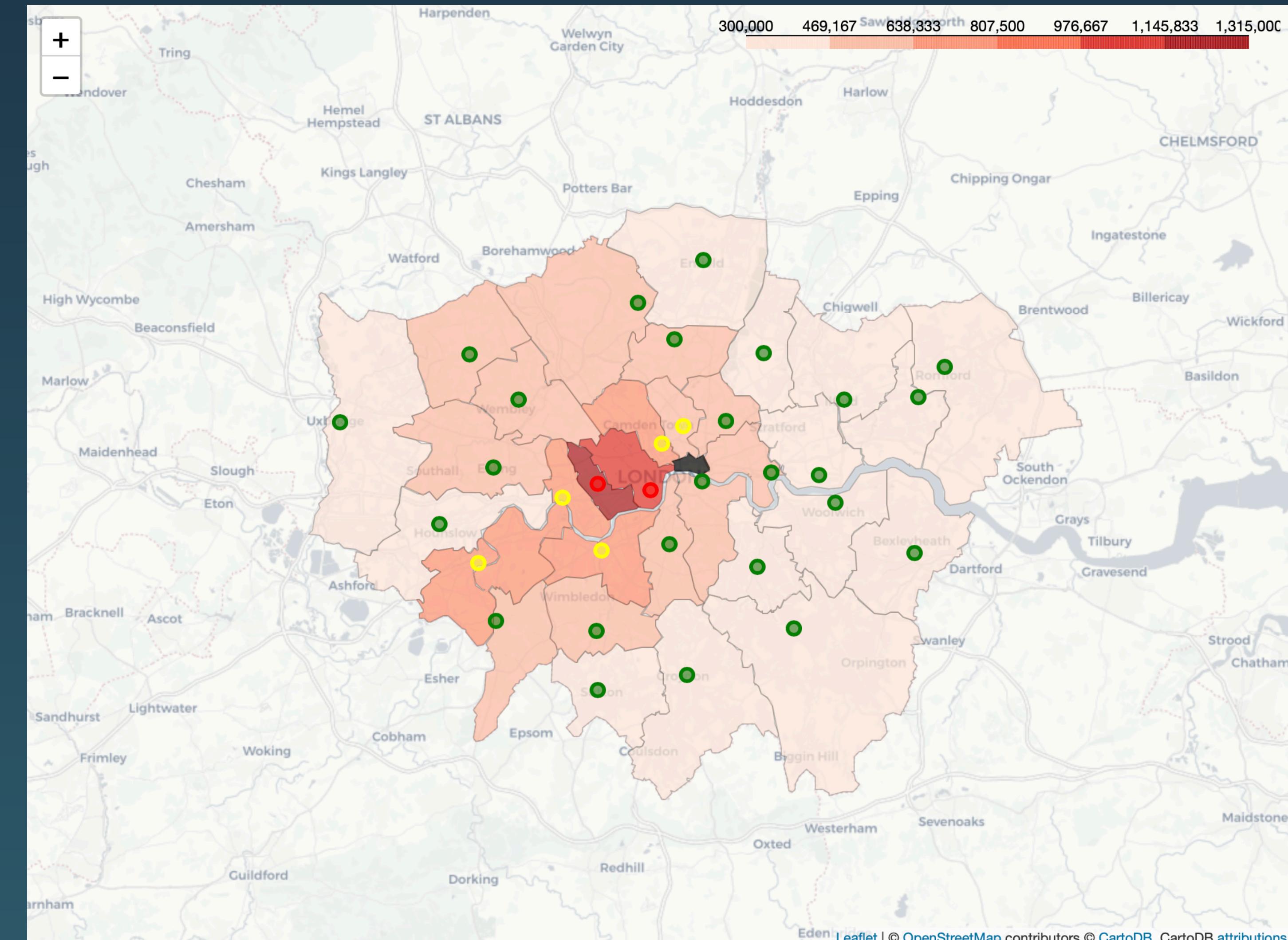
ANALYSING THE CLUSTERS

Crime and school data also showed differences between each cluster, but not to the same extent



ANALYSING THE CLUSTERS

The red cluster corresponds with the highest prices, the yellow cluster with the middle prices and the green cluster with the lowest prices



CONCLUSIONS

Borough clusters can be labeled as:

- **Green cluster** (Cluster 0) – low house prices, low air pollution, low crime rate, low number schools per population, lowest number of amenities
- **Yellow cluster** (Cluster 2) – medium house prices, medium/high air pollution, low/medium crime rate, medium number of schools per population, medium number of amenities
- **Red cluster** (Cluster 1) – high house prices, high air pollution, high crime rate, high number of schools per population, highest number of amenities

CONCLUSIONS

- Depending on an individual's priorities, each cluster may appear less or more desirable.
- Many may choose a borough in the green cluster due to the lower house prices and crime rates, but some may prefer having more amenities and schools in a borough in the yellow or red clusters.

CONCLUSIONS

- These findings could be valuable for the government to try and tackle crime and implement more affordable housing in these central boroughs.
- Furthermore, companies could utilise this data to better understand where many services are lacking and capitalise on this (e.g. opening new supermarkets), or to implement eco-friendly policies to reduce pollution in the capital.

FUTURE DIRECTIONS

Analyse data for London postcodes to find differences within boroughs

