

The Battle of Neighborhoods: move for better cultural environment?

Clustering schools in London based on house prices and the number of libraries and museums

Yifei Wang

1. Introduction

Parents choose schools for their children with great carefulness. Middle-class families have busy parents who work hard to support their families but with less time to spend with their children. Finding a school close to libraries and museums is able to provide a better environment for young students to explore the world in their spare time even when their parents are at work.



Figure 1. Boroughs in London (blue dots).

This project will, therefore, focus on the schools in London and cluster the schools into different categories according to the house prices in the corresponding boroughs (Fig. 1) and the number of cultural facilities (i.e. libraries and museums).

2. Methods

Open-source data from the website will be collected and analysed to make predictions and clustering.

(1) The mean and median house prices of each borough in 2022 will be predicted and become the features for clustering the schools in London. A linear regression model used to predict the house prices are based on the fact that London's real estate market is mature and the house prices are increasing at a relatively stable rate. Mean and median prices are both predicted to fit different custom types. For high-end needs, boroughs with large mean and median gap will be suitable. For average family, median prices will be more reliable than mean prices.

(2) Foursquare API will be used to obtain the number of libraries and museums within 1 km around each school and K-means clustering will be the method for classification. Eventually, the schools will be classified into 4 categories.

3. Data

3.1 House prices

The original dataset contains mean, median and sales prices of houses from 1995 to 2017. Mean and median prices are selected to provide a comprehensive perspective of the market to readers. The mean prices in all the London boroughs are higher than the median prices all over the period, which indicates that there were houses with better facilities or refurbishments sold at high prices. Parents are recommended to consider this situation when comparing the schools in the same borough.

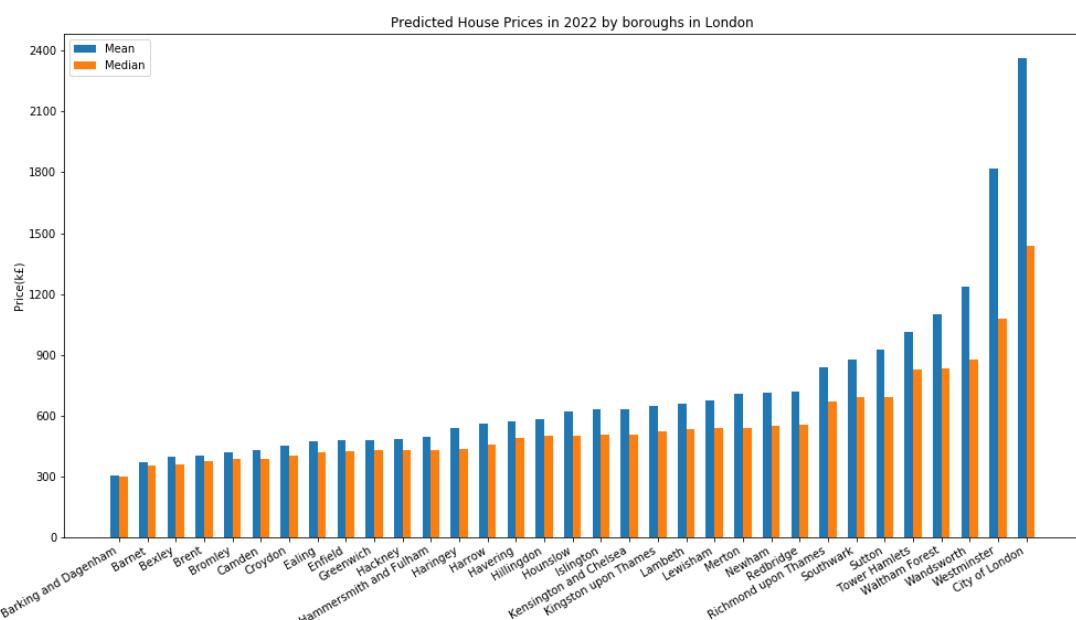


Figure 2. Predicted mean and median house prices in London boroughs and City of London in 2022.

Linear regression model was used to predict the prices in 2022. For both mean and median, the trained model showed good accuracy, with the minimum r^2 score of mean prices to be 0.861 and median prices to be 0.864. The mean and median models were then applied to predict the house prices of each borough in 2022 (Fig. 2). As shown in the figure, boroughs in the core of London (i.e. Westminster and City of London) will have the highest prices. The mean prices of these two areas are significantly higher than the other boroughs, indicating the existence of luxury houses, mostly owned by the tycoons. For middle class parents, there is no persuasive reasons to invest real estate in the centre of London if the only purpose is to find better education for the next generation.

3.2 Number of libraries and museums

Nearby venues were searched with Foursquare API to obtain the number of libraries and museums around each schools with the radius of 1 km. The number of libraries and museums was then used as features of each school for clustering. An example of the results is shown in Table. 1.

Table 1. Five schools with the number of libraries and museums counted.

School	Address	Borough	Post Code	Mean House Prices	Median House Prices	Library	Museum
Ayesha Siddiqa Girls School	165-169 The Broadway	Ealing	UB1 1LR	629923	500665	1	0
Beis Medrash Elyon	233 West Hendon Broadway	Barnet	NW9 7DG	720990	541240	1	0
Big Creative Independent School	Silver Birch House	Waltham Forest	E17 5SD	449010	433996	0	0
Wetherby Senior School	100 Marylebone Lane	Westminster	W1U 2QB	1818817	1078201	30	9
St Mary's Kilburn Church of England Primary Sc...	Quex Road	Camden	NW6 4PG	1235263	833739	3	0

3.3 Clustering

There are over 3000 schools in London. For clearer showing, only a quarter is chosen as samples and marked on the map (Fig. 3). Mean and median house prices of the borough where each school is located, and the number of libraries and museums with 1 km around each school were considered when classifying with K-Means clustering.

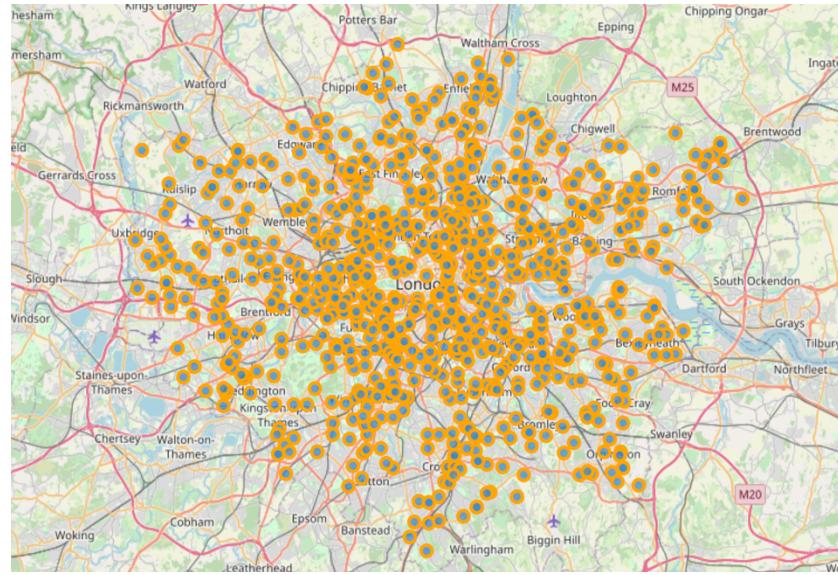


Figure. 3 Schools in London (a quarter of the total number is shown on the map).

A total of 4 clusters are identified and marker in different colours (Fig. 4). For each cluster, the average, highest and lowest median house prices are found to picture the features. The mean house prices is not considered due to the existence of luxury houses in certain boroughs which may affect the accuracy of reflecting the real circumstance.

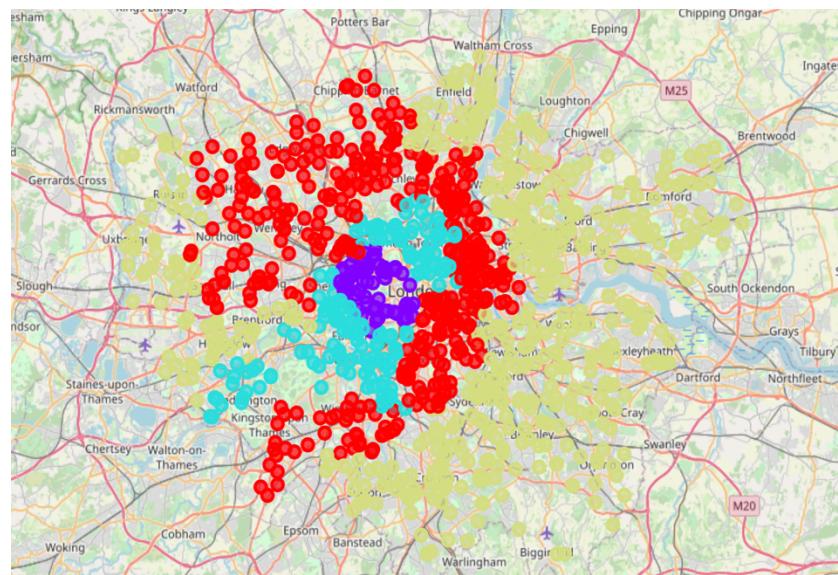


Figure 4. Clustered schools in London

4. Results and Discussion

The four clusters are ranked according to the median house prices (Table. 2). As shown in the table below, the number of libraries and museums also follow the same sequence.

Table 2. Median house prices and maximum number of libraries and museums of clusters.

	AVERAGE (£)	HIGHEST (£)	LOWEST (£)	LIBRARIES (max num.)	MUSEUMS (max num.)
PURPLE	1,253,980	1,436,269	1,078,201	50	50
BLUE	744,354	875,176	668,590	50	50
RED	525,295	554,687	491,768	27	20
GREEN	401,690	454,590	299,936	8	6

It is not surprising that schools in central London have the highest accessibility to libraries and museums (purple clusters). Clusters with higher house prices are usually surrounded by those with lower house prices (Fig. 4).

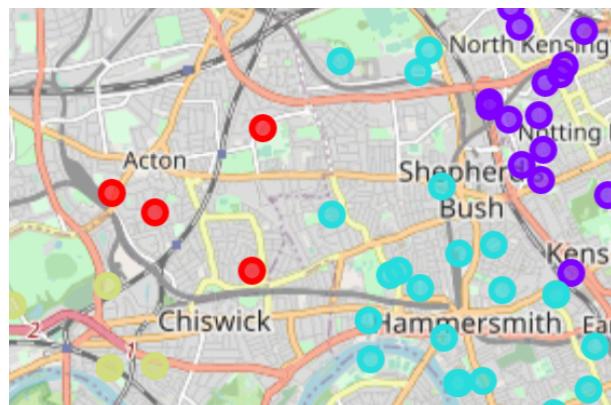


Figure. 5 Area in the west of Shepherd Bush and Hammersmith with the markers of four clusters

However, there is an area near Shepherd Bush and Hammersmith where all four clusters almost intersect together (Fig. 5). It could become a potential area for parents who are seeking for better cultural environment. House prices with different level and adjacency to center of London would be the prominent features.

5. Conclusion

Overall, 4 clusters are identified among more than 3000 schools in London. They are targeted at families with different income level. The purple clusters in the center of London ranks the

highest house prices (more than 1 million pounds) as well as the most libraries and museums around schools. The blue clusters in the outer layer of purple clusters has lower house prices of around 0.7 million pounds, but also have schools with considerable accessibility to cultural facilities. Next is the red cluster, with a moderate house prices at a level of 0.5 million. Less libraries or museums can be found around the schools in this clusters. The last clusters consist of schools mostly located in the suburban area of London with the lowest house prices and the fewest libraries/museums available.

Despite that higher house prices provide better environment, it is possible to find at least one area that is the most cost effective. Located in west London, next to Shepherd Bush and Hammersmith, the area is close to the top 2 tier with sufficient education resources. It is wise to consider the tradeoff between spending more or accessing less before making the final decision.

There are ways to optimize the results so that better description and prediction can be made. Other factors such as transportation, restaurant and supermarket near the schools and houses should be also be considered. In the end, a UI interface could be constructed showing the schools, house prices in the area and facilities availability with link to websites for house sales or school admission. This could help parents to decide if it worth moving to a new places.

Data source:

- *London house prices and School atlas :* <https://data.london.gov.uk/>
- *London borough list:* https://en.wikipedia.org/wiki/List_of_London_boroughs