Housing Purchase Analysis Based in Greater London, UK: Looking for a Suitable Community for Middle-Class Investors

Author: Kuangcheng (Aaron) ZHU

1. Introduction

London, the capital of the United Kingdom, one the financial centres of the whole world, has always been a city full of vibrant cultures, some of the world's most advanced events and plenty of superior business opportunities. Many are aiming at getting a career there, a great number of whom intend to make London their new home, but as middle-class community they cannot really afford high-end houses, and safety and convenience of living are two major factors which should also be taken under consideration. Thus, there's a need to briefly analyse the current housing market situation in London and get some advice for these ready-to-be new Londoners, with middle-class background. First, overall analysis is shown below to see why it's high time to get a real estate property.
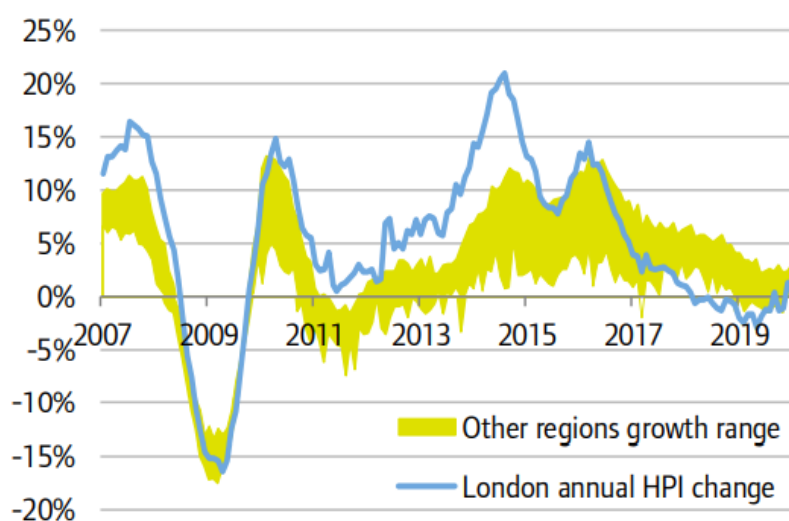

Figure 1: Annual house price change, London and other UK regions[1]

Looking into figure 1 from London House Market Report (May 2020) it's not hard to spot that from 2016 the HPI (house price index) has been going down, and since 2018, the growth rate had been negative, until the publication of this report. This situation creates a potential market opportunity for those who want to buy a property in London with limited capital, as well as the COVID-19 pandemic effect, the HPI is expected to drop sharply in a short term.
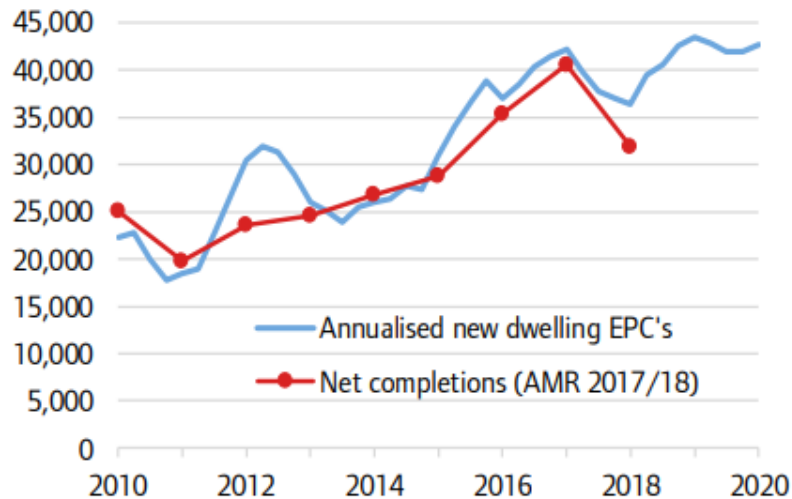
Figure 2: Energy Performance Certificates for new dwellings[1]

From the same report, the Energy Performance Certificates issued in London can be seen via figure 2. The certificate is required for all new homes so is regarded as a proxy for recent completions. Ever since 2013, more and more real estates had been guaranteed for sale and in 2019, the number of annualised EPCs reached its all-time high, as more than 43000 new homes were ready to enter the market and awaited their new owners.
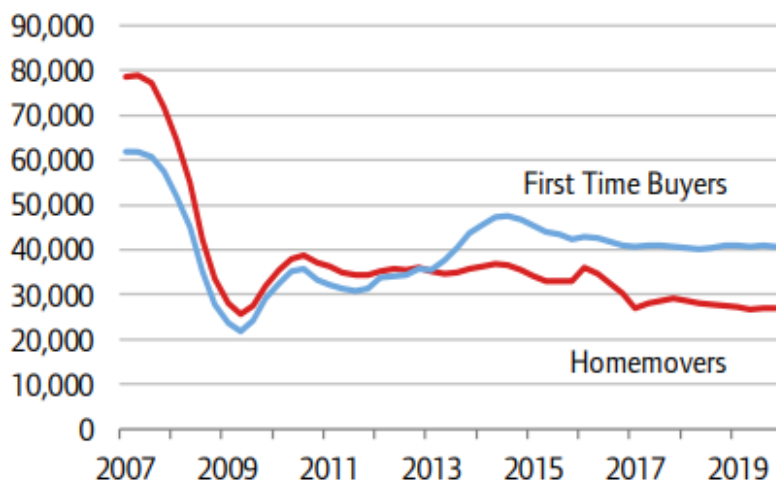

Figure 3: Annualised mortgage advances to first-time buyers and home movers[1]

Figure 3 suggests the stable level of annualised mortgage advance for first-time buyers in London. The level had dropped quite a lot compared with more than a decade ago, when financial crisis didn't yet hit the whole real estate and banking industries. Since then the level had been floating around 30000 to 50000 pounds and was around 40000 pounds in early 2020.

2. Data Description

All the raw data used are from London Datastore (a data site set by London government), Stuart Grange's GitHub page and Foursquare (a site which can provide geo locations via APIs).

Some of the major data describing the London house market are from *London Datastore*, where I get these four sets of data: 1. *Number and Density of Dwellings by Borough*[2]; 2. *Recorded Crime: Borough Rates*[3]; 3. *Vacant Dwellings*[4]; 4. *Average House Prices by Borough, Ward, MSOA & LSOA*[5]. The first dataset is used to show the density of dwellings of each borough. The second dataset is used to show the crime rate of each borough. The third dataset is used to show the number of vacant dwellings around London and the number is seen as a proxy of dwellings on sale, whether newly developed first-hand dwellings or second-hand dwellings. The last dataset is used to show the house prices in London, categorised by borough, and I also get the ward data under each borough.

From Stuart Grange's Home[6], I get the GeoJSON file which includes all the boundary coordinates of all the London boroughs. I later use this JSON file to draw the boundaries of boroughs on the map, thus London can be divided into boroughs and can be seen on the map clearly.

By using Foursquare I get call the APIs to get the information of some of the most visited venues around a certain place. I upload the coordinates of a place (for example, a ward) and Foursquare APIs will return me a JSON file which contains the info of popular venues around the place (ward), like coordinates, venue type and visited frequency.

Furthermore, to keep consistency of time period across different datasets, all the data of year 2017 are used for later analysis to avoid missing data problem. During the 3 years between 2017 and 2020, there wasn't any surging development in any borough on real estate, so we assume the absolute values are not that important, and later analysis will focus on relative comparison between boroughs.

3.   Methodology and Discussion

Firstly, data cleaning and reconstruction are executed so they can be directly used for data analysis, and the Pandas library is used in this process. After completing this process, with the help of the Folium library, some maps are generated to more intuitively view the features of the London house market and London boroughs.
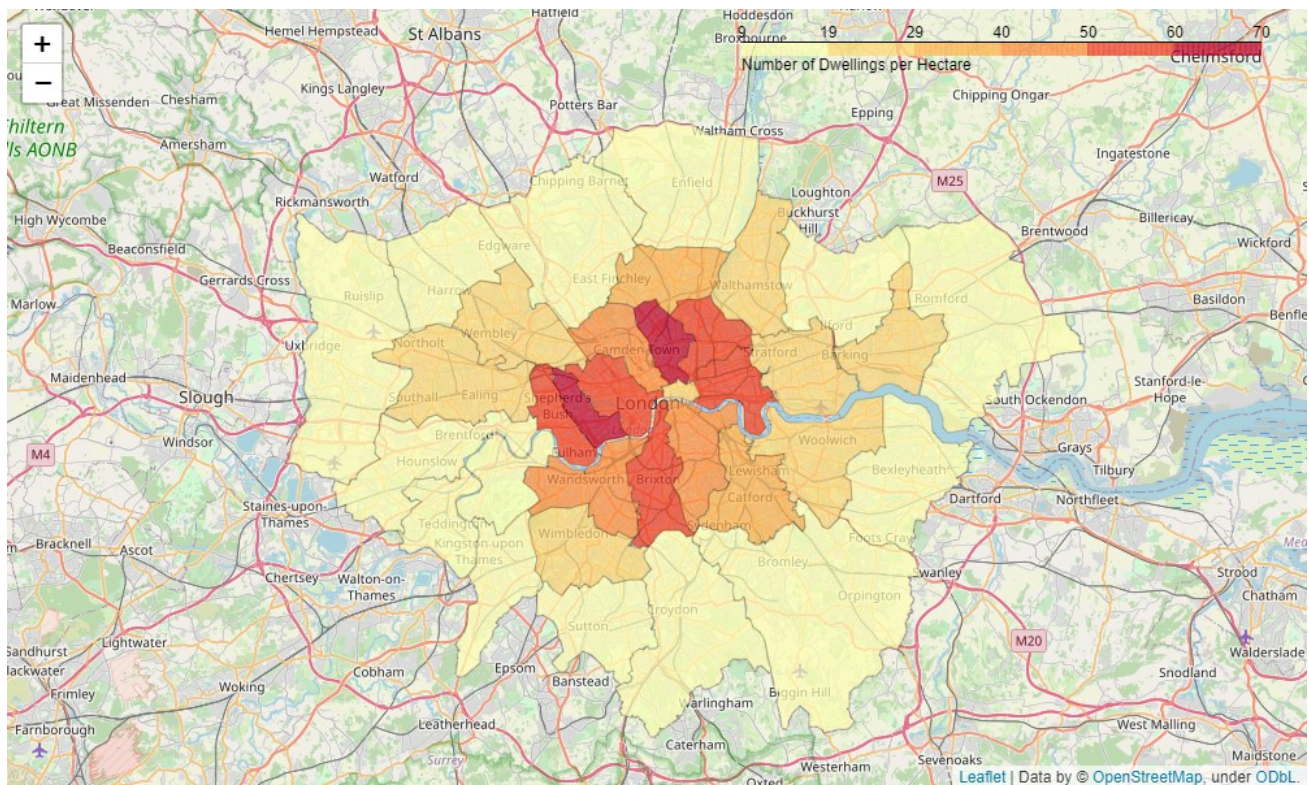

Figure 4: Map of dwelling density by borough

Figures 4 gives an idea of how close dwellings are settled around different boroughs in London. Islington in central-north London, together with Kensington and Chelsea in central-west London, are two boroughs with highest residence density. 19 out of 33 boroughs have a density larger than 40 dwellings per hectare, which means these boroughs are good choices of middle-classers, not only because these boroughs have a better selection on transportation and commuting (usually populated areas get more bus stops, underground/overground and national rail stations), but also because these boroughs are closer to city centre, where people go to work.
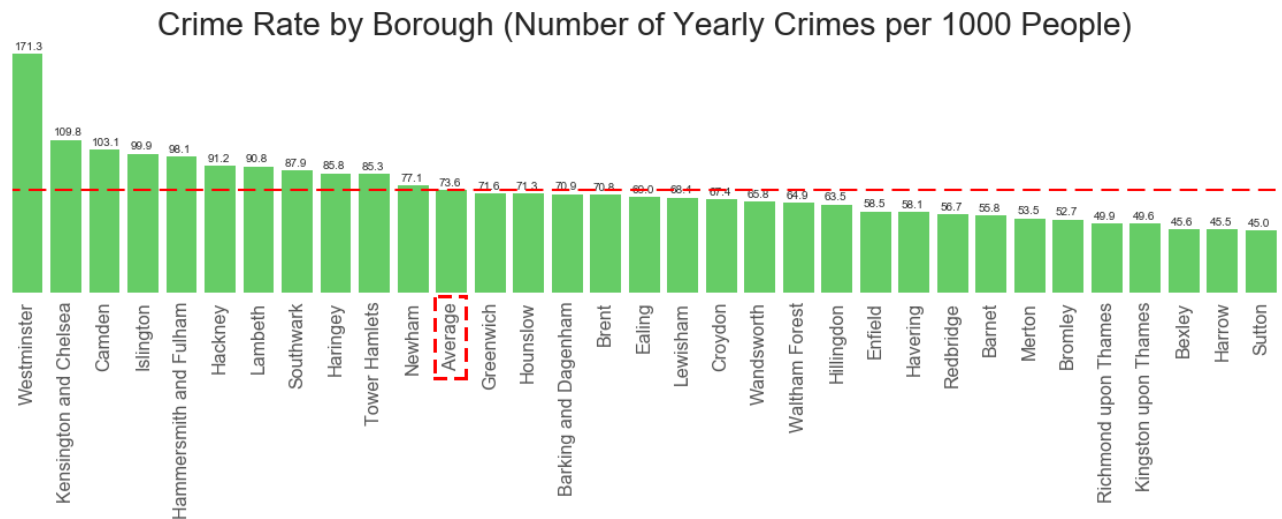
Figure 5: Bar chart of crime rate by borough

Crime rate is an important indicator for safety and the crime rate of each borough is displayed on the above figure. If we set the standard that we don't want boroughs with a crime rate greater than average crime rate of all boroughs, then we drop the top 11 boroughs with high crime rate.
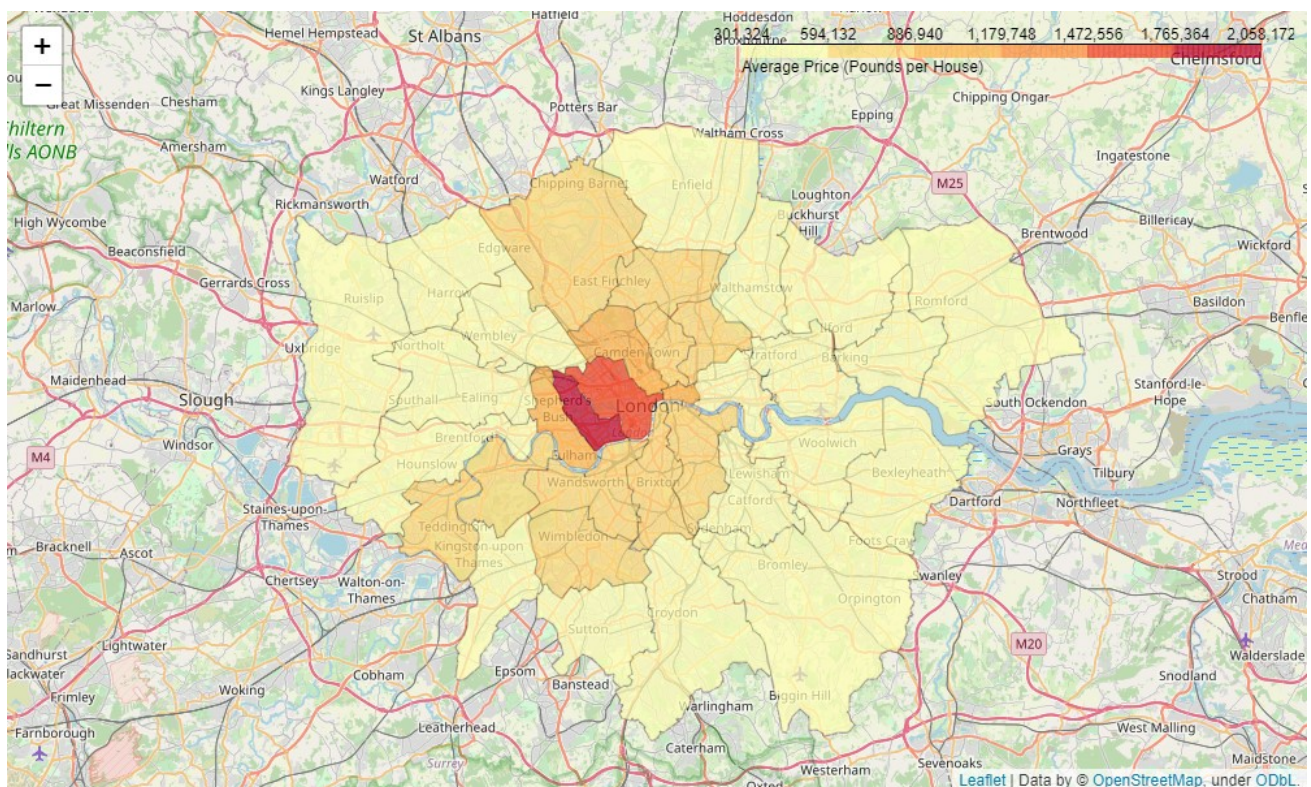

Figure 6: Map of average house price by borough

In this house price map, it's easy to spot that Kensington and Chelsea and Westminster are the boroughs of highest average house price, followed by central boroughs Hammersmith and Fulham, Islington and City. Boroughs north and south of these boroughs are of medium price level.
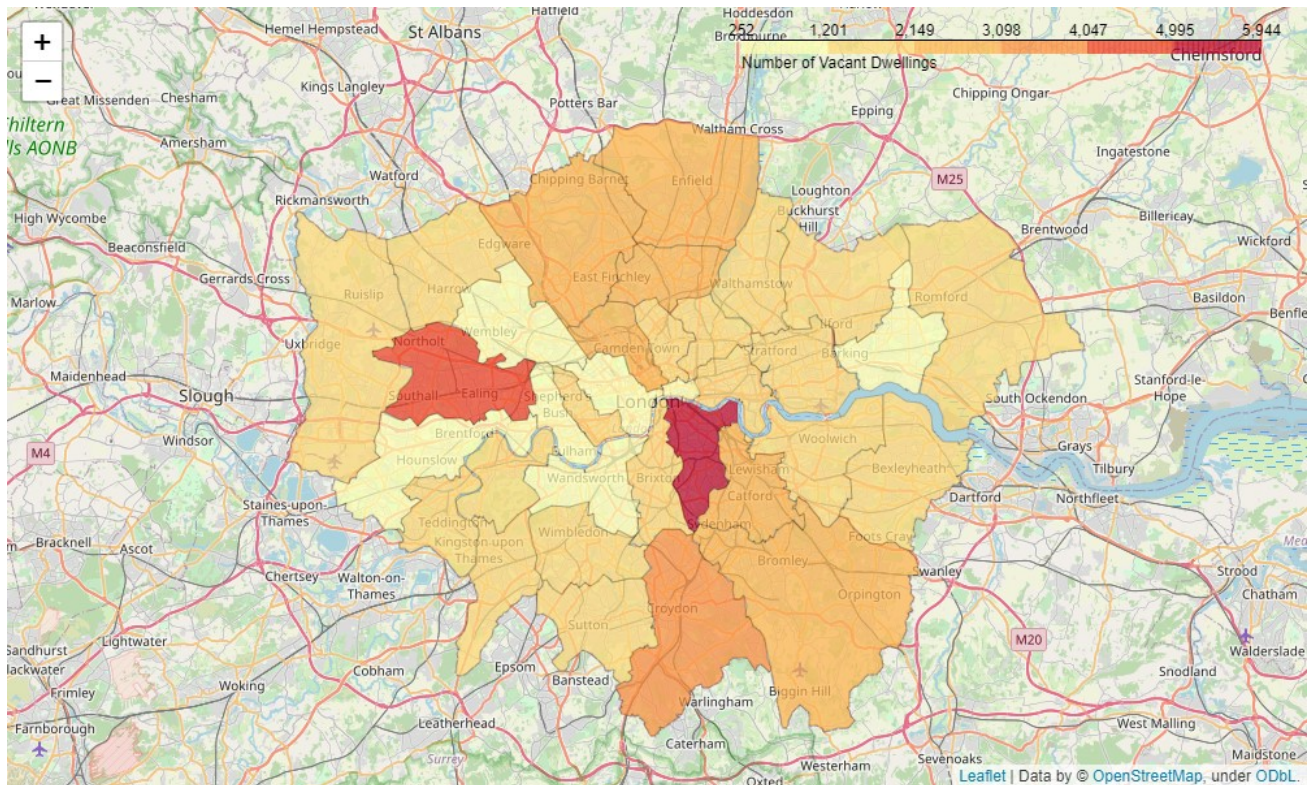
Figure 7: Map of vacant dwellings by borough

As an indicator and proxy of available houses, the number of vacant dwellings can give guidance on which borough can have more choices when it comes to deeper checking. Within a certain borough, if there're more vacant houses, the possibility of finding somewhere suitable can be larger, apparently. There're some boroughs (Hounslow, Brent, Hammersmith and Fulham, Wandsworth, Westminster, City and Barking and Dagenham) with fairly limited numbers of available residences and they might be out of consideration due to the lack of choices.

After going through all the features of boroughs we gradually set up our standards of making choice. Based on such criterion: 1. The density is (number of dwellings per hectare) larger than 29; 2. The crime rate is under the average level; 3. The price level is not high. Now we can find two boroughs left for further investigation: Wandsworth (probably with fewer available housing options) and Southwark, both south of the River Thames.
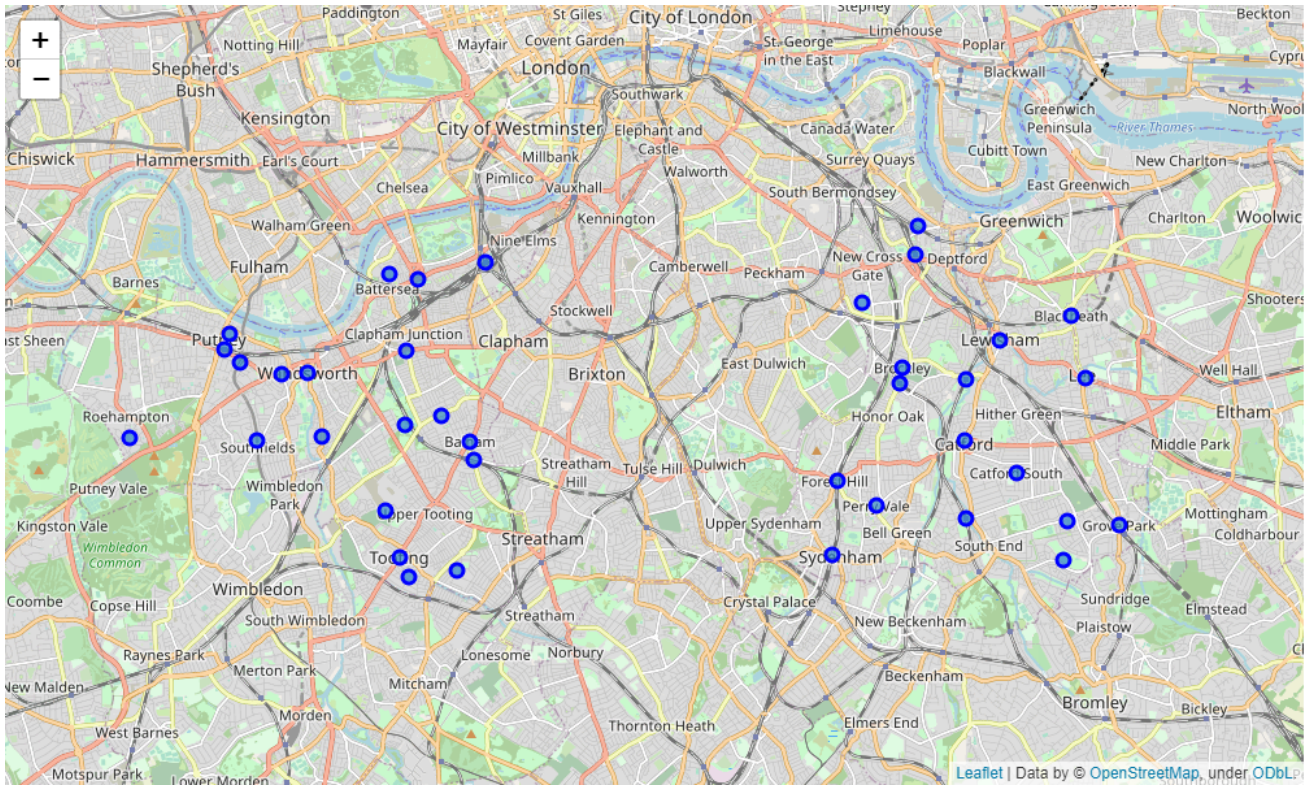
Figure 8: Map of wards in Wandsworth and Southwark

With the help of GeoPy library, I get the coordinates of all the wards under the two selected boroughs: Wandsworth (the left cluster) and Southwark (the right cluster). Thanks to the Folium library I am able to put the locations of wards on the map and each maker represents a ward.

The next step is using Foursquare APIs to get venues around all the wards. The coordinates of top 100 venues within 500 metres of each ward are sorted out, with venue name, venue category and visited frequency. Later, by utilising the machine learning library Scikit-learn, K-Means clustering process is executed on these venues to generte an overall look on how the neighbourhood of a ward differ from others' neighbourhoods.
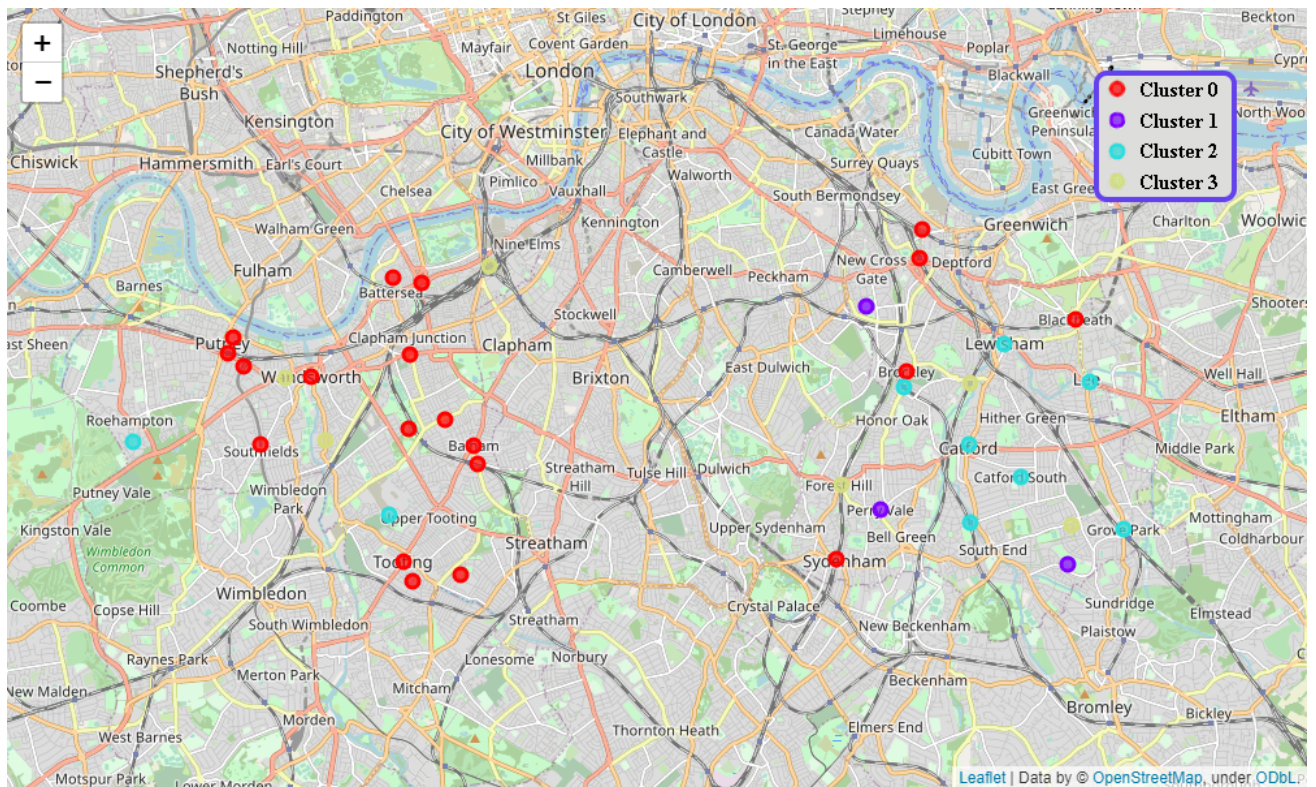
Figure 9: Map of clustered wards in Wandsworth and Southwark

After the K-Means clustering algorithm is executed the 38 wards under the two boroughs, Wandsworth and Southwark, are distributed into 4 clusters based on their nearby venue similarity. The 4 clusters are shown on figure 9 with different colours.
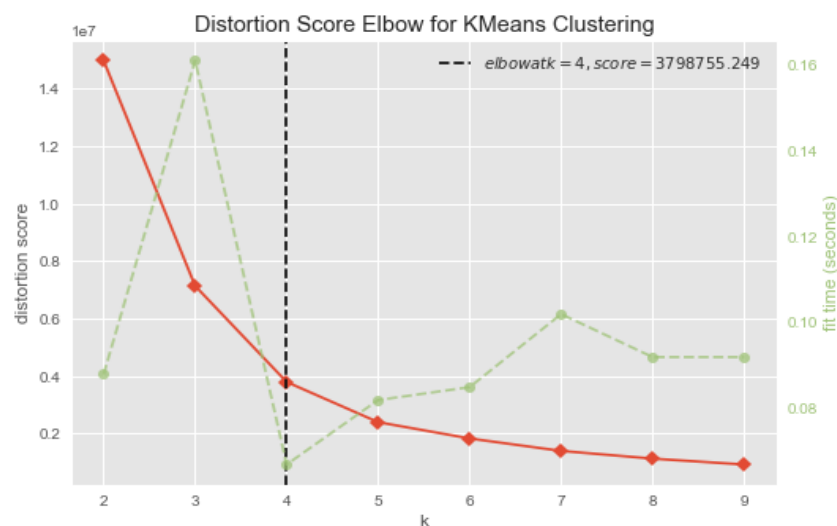

Figure 10: Test results for K-Means clustering

By getting the distortion score elbow via the Yellowbrick library, we can justify why there're 4 clusters. The results suggest an optimal k with the value of 4 should be used for clustering. But how on earth are these 4 clusters different from each other?
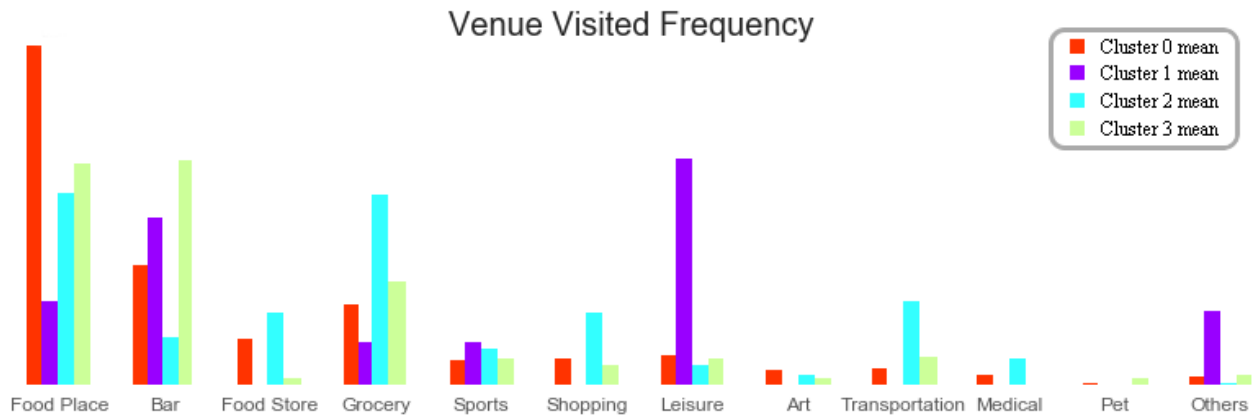
Figure 11: Comparable info of different clusters

From figure 11, food place (restaurant) is the most welcomed venues around cluster 0 wards, bar and grocery also consist a very big part of their nearby venues. For cluster 1 wards, leisure area occupies the biggest part of visits, while bar is popular as well. For cluster 2 wards, food place selection and grocery store selection can be plenty, but bar selection is poor. For cluster 3 wards, food place and bar are popular and might be all around nearby. Overall, cluster 0 and 3 wards seem to be the most alive wards, with their busy restaurants and bars, cluster 2 wards are great choices for grocery shopping and cluster 1 wards are the places for people who like parks and nature.
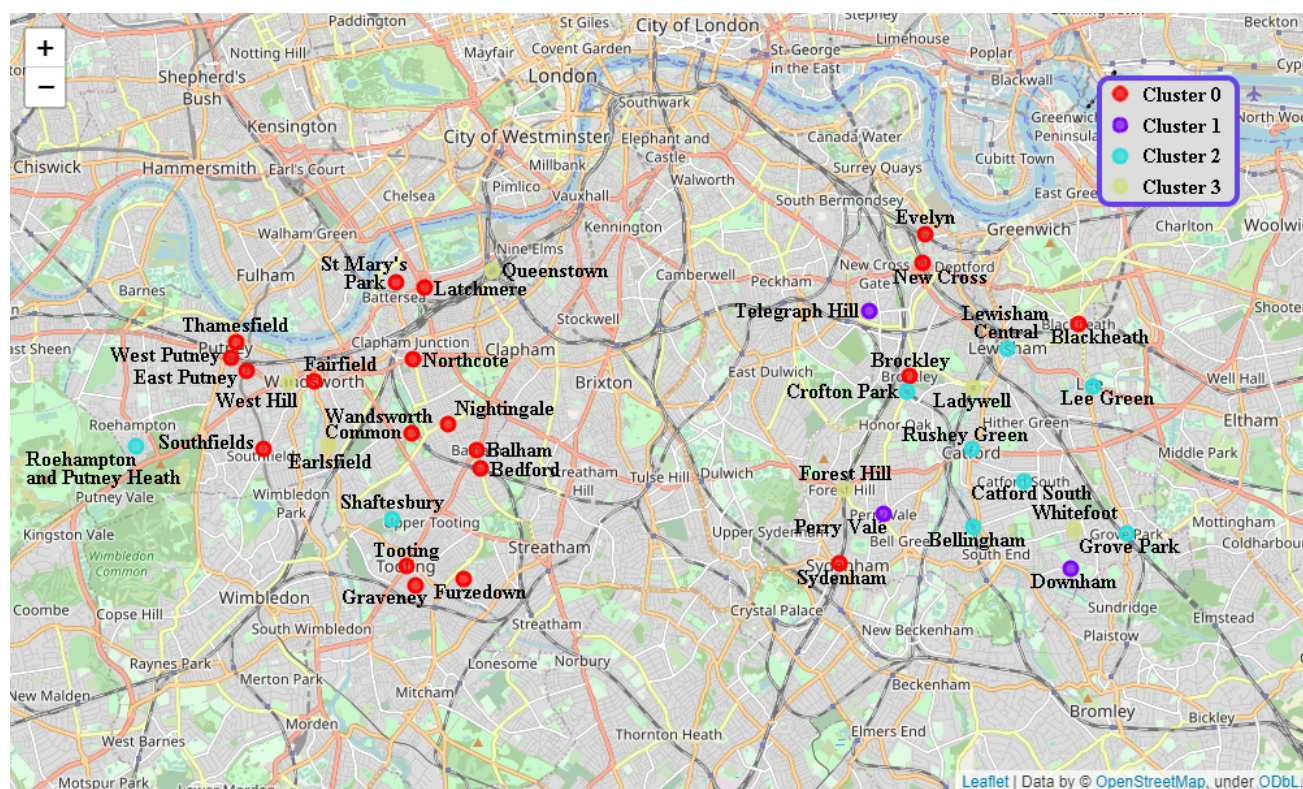
4. Results and Recommendation

Figure 12: Map of clustered wards with names

The overall results are displayed above as a map with ward names. It's not hard to find out that there're 4 clusters of different types of wards in 2 recommended boroughs (Wandsworth and Southwark).

The 2 recommended boroughs are of relatively low crime rate, and their residential areas are populated and have medium house price among inner London. Thus, the 2 boroughs are safe, easy-to-commute and fairly affordable for middle-classers who have a plan to settle down in London, especially Southwark because of its larger number of vacant dwellings and its bigger possibility to find a suitable house among available residences. Within the 2 boroughs, there're 4 types of wards, and depending on every buyer's lifestyle, one can choose his or her own preferred area to seek for a property. As discussed before, cluster 0 and 3 wards seem to be the most alive wards, with their busy restaurants and bars, cluster 2 wards are great choices for grocery shopping and cluster 1 wards are the places for people who like parks and nature.

5. Conclusion

In conclusion, this article provides some insights on what consideration middle-class people should take and what options they can have, if they prepare to settle down in London. According to the article, based on historic data analysis as well as the newest trending, in the foreseeable future, the housing market in London will offer a superior opportunity for investors, and for middle-class investors among them, getting a place which is safe, easy-to-commute and affordable would be their main consideration of spending money on a real estate property. Wandsworth and Southwark are two boroughs highly recommended by this article, for they both satisfy all the three requirements we raised above. And within Wandsworth and Southwark, different wards are separated into 4 clusters depending on their nearby venues and their functions. It is highly recommended that investors have their own opinions, at the same time, relying on a variety of housing purchase analysis like this article to make judgment, therefore, they can make the right decision for themselves.

Reference and Data Sources

[1] GLA Housing and Land: *London Housing Market Report (May 2020)*.
     https://data.london.gov.uk/housing/housing-market-report/

[2] Ministry of Housing, Communities & Local Government (MHCLG): *Number and Density of Dwellings by Borough*. https://data.london.gov.uk/dataset/number-and-density-of-dwellings-by-borough

[3] Metropolitan Police Service: *Recorded Crime: Borough Rates*.
     https://data.london.gov.uk/dataset/recorded_crime_rates

[4] Ministry of Housing, Communities & Local Government (MHCLG): *Vacant Dwellings*.
     https://data.london.gov.uk/dataset/vacant-dwellings

[5] Land Registry: *Average House Prices by Borough, Ward, MSOA & LSOA*.
     https://data.london.gov.uk/dataset/average-house-prices

[6] Stuart Grange's Home. https://skgrange.github.io/data.html