**Where to buy as a first-time buyer in London**

El Ghali Alaoui

January 6, 2020

**1. Introduction and business problem**

London Housing Market is famous for being one of the most expensive of the world with only a small portion of first-time buyers (FTBs) being able to afford a house and get on the property ladder.

London is administratively divided into 32 boroughs, with significant disparity between them in terms of average house price and venues.

In terms of FTBs profile, the average age to purchase a house for the first time in London is 33 years old. First-Time buyers are mostly couples buying together (77%). 37% of FTBs have children.

The average house price in London purchased by first time buyers is 400 K.

The problem this analysis aims to solve is where (i.e. which boroughs) should first-time buyers purchase a house in London based on affordability (i.e. average house price) and infrastructure (i.e. venues) criteria.

**2. Data acquisition and cleaning**

**2.1 Data acquisition**

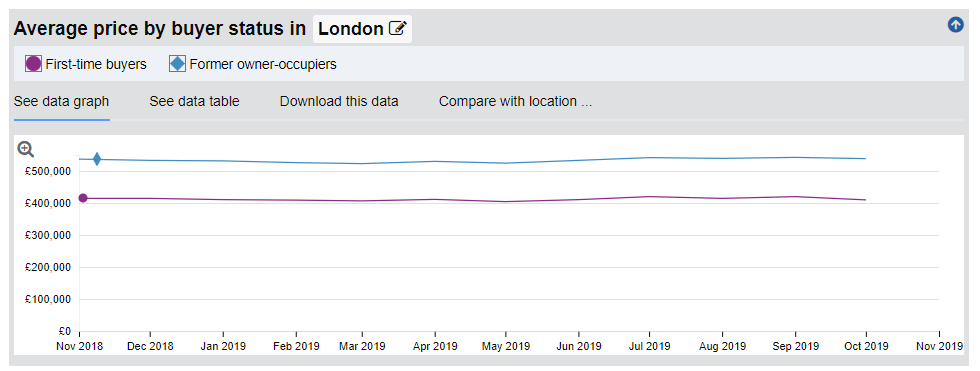
Data used to solve the problem comes from various sources.

First time buyers profile information can be found in the link below:

<https://www.estateagenttoday.co.uk/features/2019/11/the-changing-profile-of-first-time-buyers>

Average house price purchased by first time buyers is sourced from government and land registry statistics:

<https://landregistry.data.gov.uk/app/ukhpi/browse?from=2018-11-01&location=http%3A%2F%2Flandregistry.data.gov.uk%2Fid%2Fregion%2Flondon&to=2019-11-01>



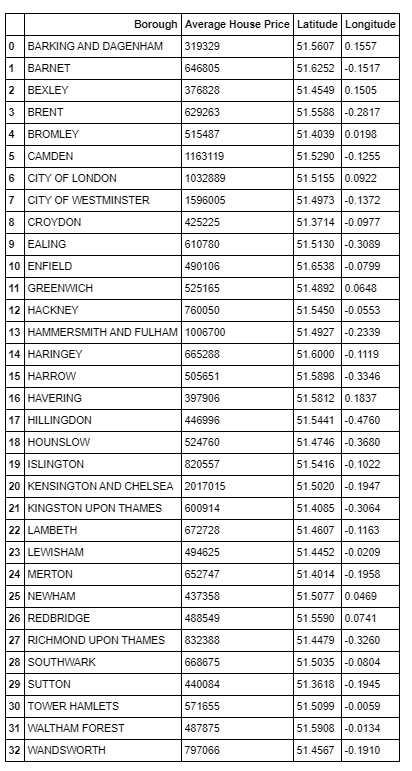
I have used HM land registry standard report service to generate a report for average house price per borough (please see <https://landregistry.data.gov.uk/app/standard-reports> ) and uploaded the generated report on my GitHub repository: <https://github.com/Elghalilou/Capstone/blob/master/avgPrice-county-GREATER_LONDON-by-district-any-2019-Q3.xlsx>

Borough coordinates can be found on Wikipedia: <https://en.wikipedia.org/wiki/List_of_London_boroughs> . I have populated the data in a csv file uploaded on my GitHub repository: <https://github.com/Elghalilou/Capstone/blob/master/Latitudes%20and%20longitudes%20csv%204.csv>

I have only used the average house price per borough and coordinates data for the analysis, the first-time buyers profile data and average house price purchased by FTB’s have been used to make recommendations.

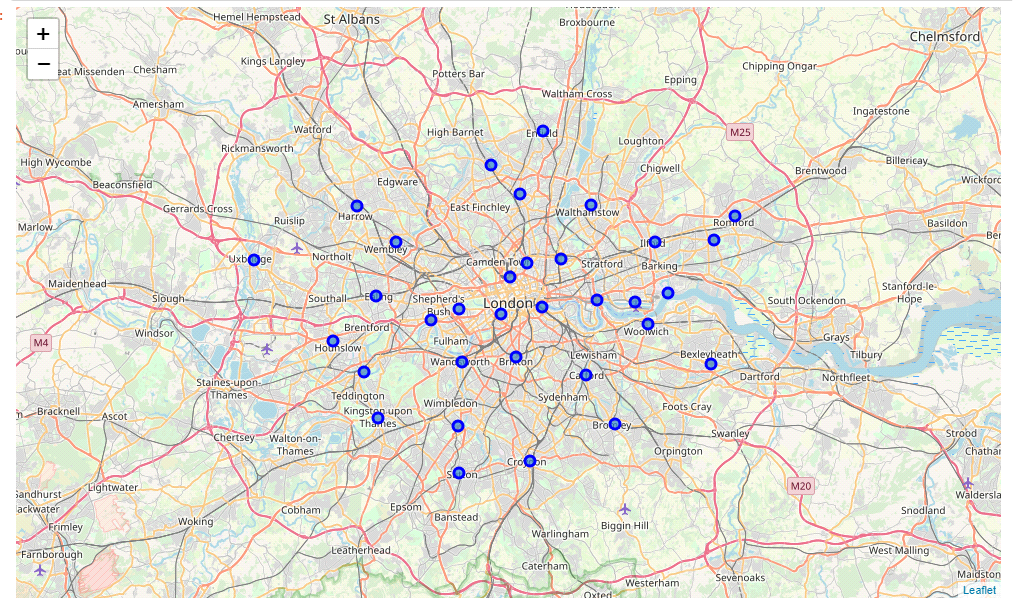
**2.2 Data cleaning**

I have used panda read method to convert average house price per borough data and borough coordinates data into panda data frames, cleaned the data by dropping irrelevant rows and columns, and merged the two data frames into a single data frame with coordinates and average house price per borough as per below.



**3. Methodology:**

I have used geopy library to get the latitude and longitude values of London and used boroughs coordinates data as well as python folium library to create a map of London with boroughs superimposed on top.



I then used four square API to explore each borough and get the 100 top venues for each borough

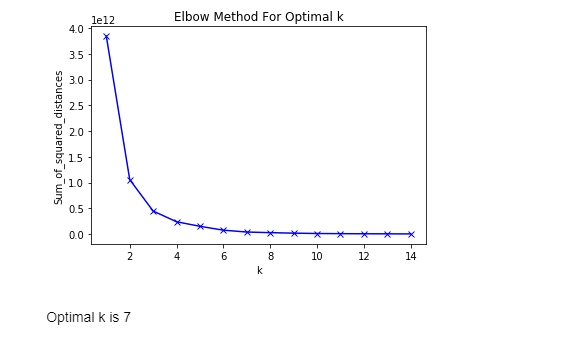


I have used one hot encoding to convert categorical data into numerical values, grouped data by boroughs and the mean of the frequency of occurrence of each venue category per borough, and added average house price to create a new data frame with frequency of occurrence and average house price per borough.

*(see Notebook 'El Ghali Assignment 11' published on my GitHub repository for data frame created as it is too large to be included in the report –Name of data frame is ‘London\_grouped\_avg')*

As discussed in the intro, the purpose of the analysis is to recommend which boroughs should FTBs buy the first home in based on venues and average house price criteria, hence I have used k-means algorithm on the data in this new data frame to segment data and cluster the boroughs based on average house price and venue categories.

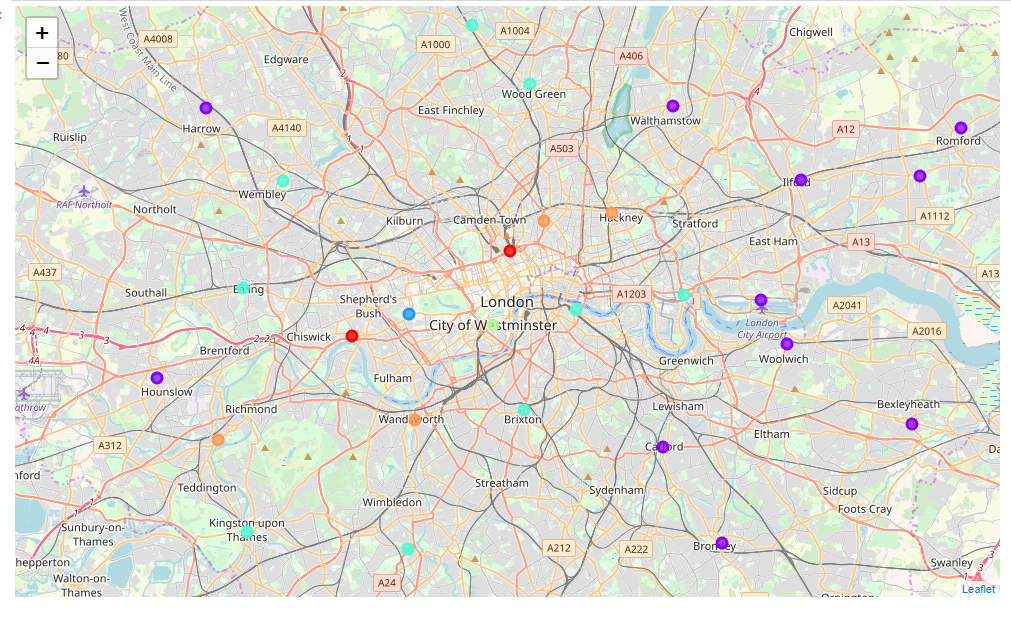
Before running k- means on the data, I have used elbow method to determine optimal k,



Based on the elbow method, the optimal k is 7, hence I have run k-means on the data to cluster the borough in 7 clusters and created a new data frame that includes the cluster labels as well as the average house price and top 10 venues for each borough

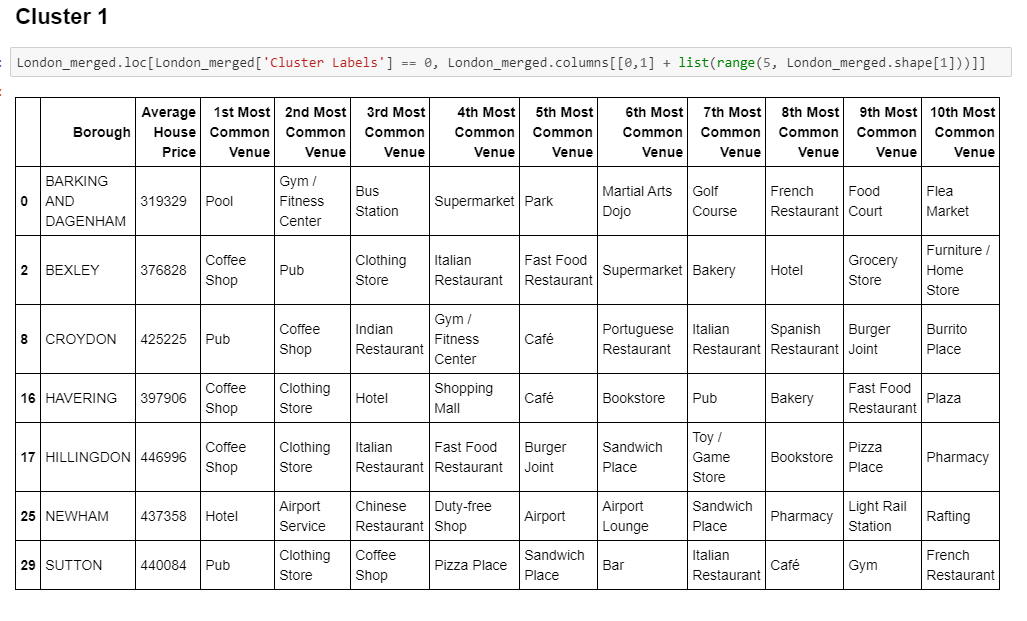
*(see Notebook 'El Ghali Assignment 11' published on my GitHub repository for data frame created as it is too large to be included in the report –Name of data frame is ‘London\_merged)*

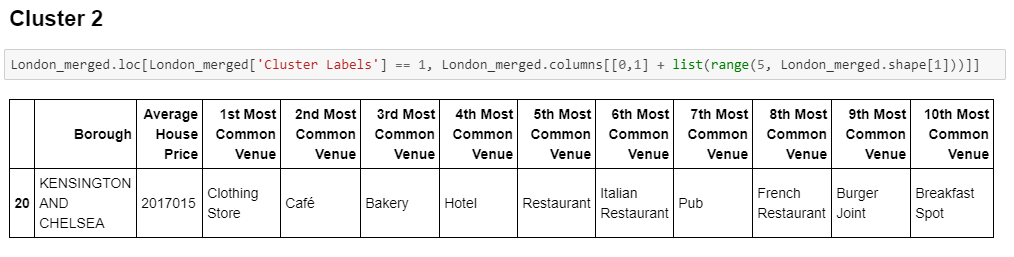
I have Used python library to visualize clusters in a map

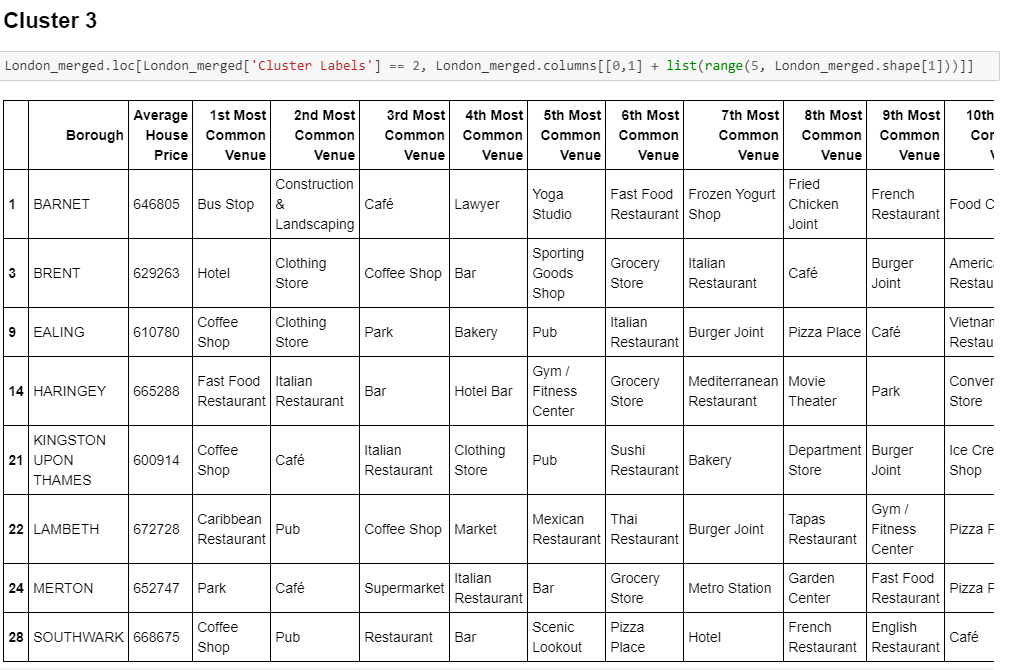


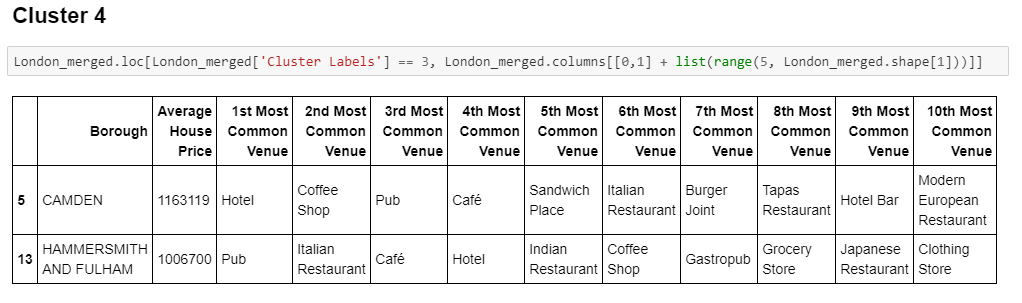
**4. Results:**

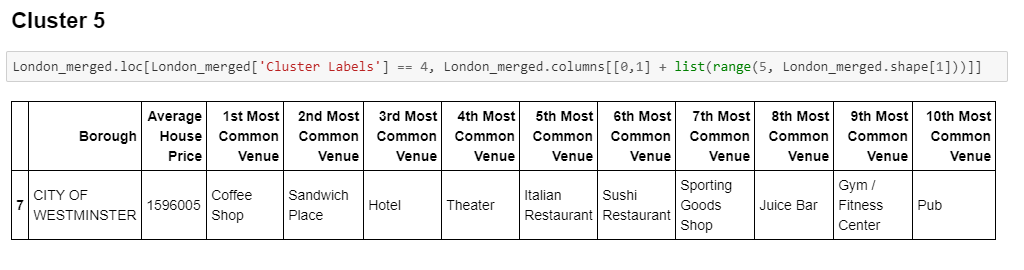
Let's examine each cluster

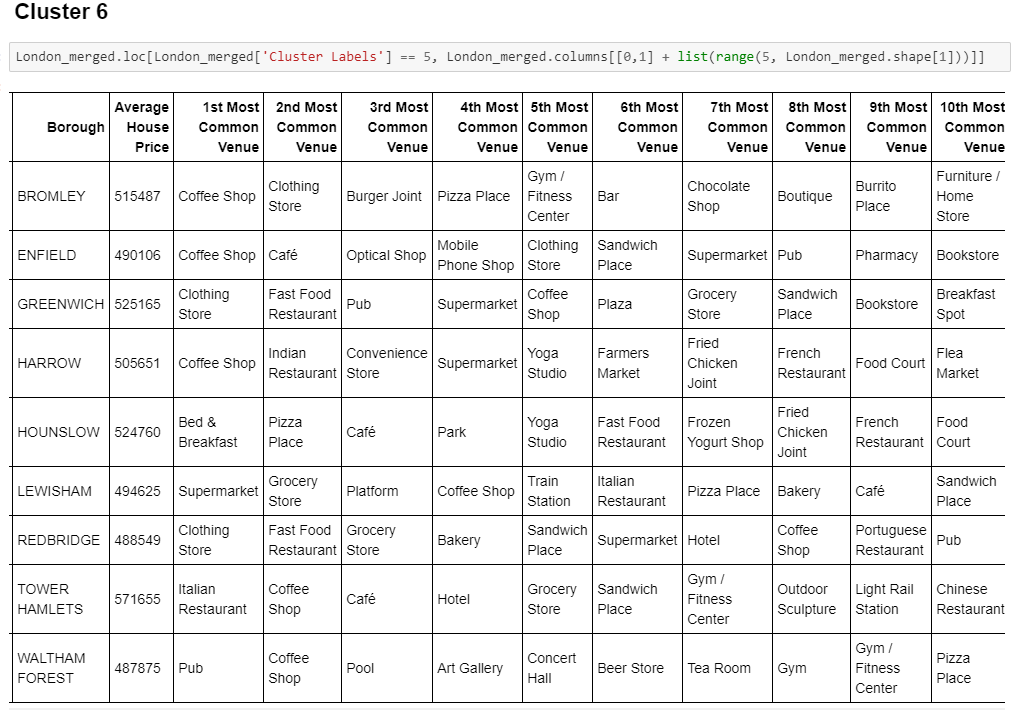


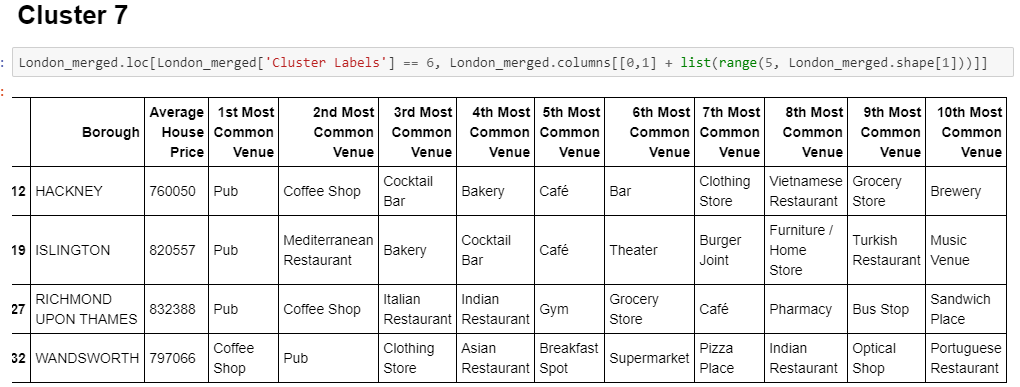












**6. Discussion**

Based on the results, we can immediately see that boroughs in clusters 2,4 and 5 are very expensive with average prices 2 to 5 times higher than the average purchase price for FTBs (400K) and are likely to be out of touch for FTBs.

Boroughs in cluster 1 is are in line with the average house price purchased by FTBs (400K) and perfectly fit affordability criteria for FTBs.

Borough in clusters 6 (between approx. 500 - 600K), 3 (between approx. 600K-700k) and 7 (between 700K and 800K) are gradually higher than the average purchase price purchased by FTBs and potentially affordable for a small portion of FTBs with better incomes.

Based on national average house price for first time buyers we can safely assume that only boroughs in clusters are 1,3,6, 7 are likely to be affordable for FTBs and should be focused on when looking to purchase a first house.

Now let’s examine the most common venues in each of these clusters.

If we regroup the most common venues per category, we obtain the following table for each cluster.

Cluster 1

|  |  |
| --- | --- |
| Most Common Venue Category | Occurrence in % |
| Drinking and Dining (Restaurants, bars, pubs, coffee shops) | 50% |
| Leisure/entertainment (pool, gym, park…) | 16% |
| Shops (Clothing store, shopping mall,) | 15% |
| Other (Bus station, airport lounge, plaza…) | 8% |
| Groceries/Supermarkets | 7% |
| Health and Services (Pharmacy, Optical shop…) | 4% |

Cluster 6

|  |  |
| --- | --- |
| Most Common Venue Category | Occurrence in % |
| Drinking and Dining (Restaurants, bars, pubs, coffee shops) | 48% |
| Shops (Clothing store, shopping mall,…) | 16% |
| Leisure/entertainment (pool, gym, park…) | 13% |
| Other (Bus station, airport lounge, plaza…) | 11% |
| Groceries/Supermarkets | 8% |
| Health and Services (Pharmacy, Optical shop…) | 4% |

Cluster 3

|  |  |
| --- | --- |
| Most Common Venue Category | Occurrence in % |
| Drinking and Dining (Restaurants, bars, pubs, coffee shops) | 70% |
| Leisure/entertainment (pool, gym, park…): | 11% |
| Other (Bus station, airport lounge, plaza…) | 8% |
| Shops (Clothing store, shopping mall,…) | 6% |
| Groceries/Supermarkets | 5% |
| Health and Services (Pharmacy, Optical shop…) | 0% |

Cluster 7

|  |  |
| --- | --- |
| Most Common Venue Category | Occurrence in % |
| Drinking and Dining (Restaurants, bars, pubs, coffee shops) | 68% |
| Leisure/entertainment (pool, gym, park…): | 11% |
| Shops (Clothing store, shopping mall,…) | 8% |
| Groceries/Supermarkets | 5% |
| Health and Services (Pharmacy, Optical shop…) | 5% |
| Other (Bus station, airport lounge, plazza…) | 3% |

We can see that although all clusters the most represented categories is drinking and dining, categories are more evenly distributed in cluster 1 and 6, while in cluster 3 and 7, dining and drinking is overly represented.

Also, there's more variety and balance in clusters 1 and 6 with more local amenities, health and services, leisure and shops and a great variety of drinking and dining options.

If we take into account FTBs profile data discussed in the intro (most likely to be in their thirties, buying as a couple, and potentially with a child) boroughs clusters 1 and 6 are more likely to suit FTBs profile, while clusters 3 and 7 with over representation of drinking and dining venues are more likely to suit a student/single/renting lifestyle.

**7. Conclusion**

Based on the above, if we consider both affordability and venue criteria as well as FTBs profiles, the best boroughs to invest in as FTB are first boroughs in cluster 1 and potentially borough in cluster 6 for FTBs with higher income.

Boroughs in cluster 3 and 7 are less likely to suit FTBs given their higher average price and venue categories more suited for a student/single/renting lifestyle.

Boroughs in cluster 2,4,5 are very likely to out of touch for FTBs as too expensive.