**Housing Prices & Venues Data Analysis of London**

Priyanka Choudhary

March 10, 2021

**1. Introduction**

**1.1 Background**

From its status as a global capital market to a leading centre for history, culture and education, London’s appeal shows little sign of abating. A web of 33 thriving boroughs, the capital is an incongruous mix of old and new, a city where the cutting-edge and uber-cool innovation hubs stand effortlessly alongside historic and iconic landmarks.

Investment across the breadth of the City continues apace, its hotspots continuously reinvented and once-forgotten corners regenerated. Well-connected within London, the UK and beyond, the capital is always accessible, and the impending Elizabeth Line will tighten links between east to west even further. Visitors can take their pick from four world heritage sites, and with a host of parks, museums and galleries, even the locals will struggle to get bored.

**1.2 Problem**

The residents across the breadth of the boroughs cited infrastructure, amenities, schools and venues are key attributes that drew them to an area. But the over-riding factor for many is the people within their reach, who, it’s widely accepted help make their locality a community, and their property a home.

**1.3 Interest**

The Data that will contribute in determining the qualitative aspects of a locality/borough like the various schools, offices and restaurants available at a nearby distance which can help a potential resident make a decision based on his needs and requirements and priorities. This project will also help a buyer/renter analyse the property price trend across the 33 boroughs in London for them to decide on the affordable borough.

**2. Data acquisition and cleaning**

**2.1 Data sources**

The descriptive data for London boroughs can be found on Kaggle. For my analysis I have used 3 datasets from Kaggle.

First dataset has population, area, distance from centre and the geospatial data for latitude and longitude of all 33 boroughs in London. I have extracted the columns for latitude and longitude from this dataset as the population data is not required for my analysis.

Second dataset contains the housing price data, number of houses sold and number of crimes for past years for each borough in London. I have extracted the data for past 3 years from this dataset and calculated the average price for same.

Third dataset contains the borough to postal code mapping and the location (inner/outer) data for each borough in London. I have used this dataset for the purpose of combining the above 2 datasets and exploring London on the inner or outer parts.

**2.2 Data Cleaning**

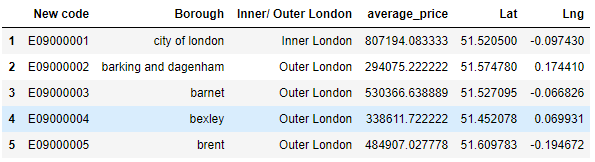
Data downloaded or scraped from multiple sources were combined into one table. There was a lot of missing values which were eventually removed from the final dataset as they did not add any value to the analysis. The data was also filtered to analyse the data for past 3 years only. Any duplicate entries were removed.

**2.3 Feature selection**

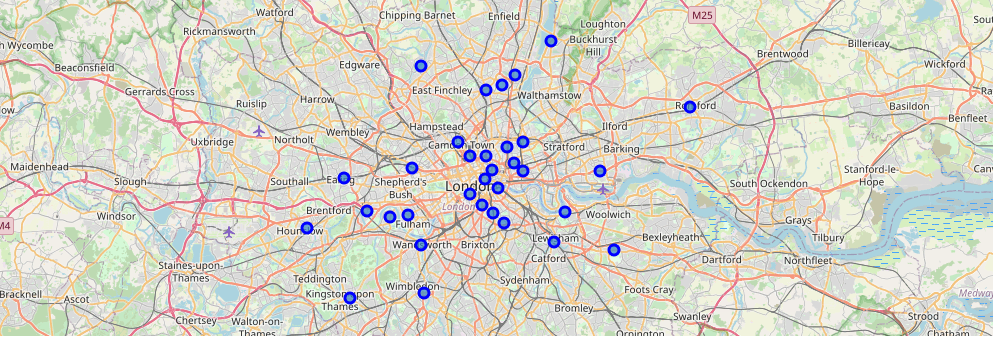
For the purpose of property price evaluation, the average price of properties for past 3 years was considered for each borough. For venue exploration Foursquare API was used to get the nearby venues based on the latitude and longitudes of each borough.

**3. Methodology**

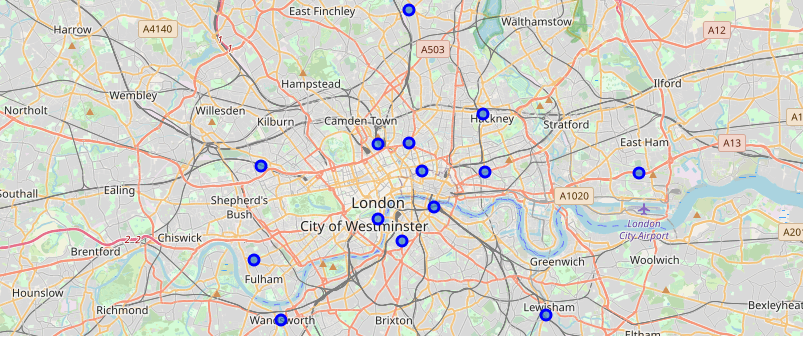
As a database, I used GitHub repository in my study. My master data which has the main components Borough, Average House Price, Latitude and Longitude and postal code informations of the city.



I used python folium library to visualize geographic details of London and its boroughs and I created a map of London with boroughs superimposed on top. I used latitude and longitude values to get the visual as below:

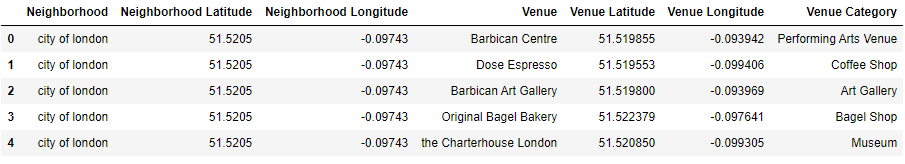


*Figure: All 33 Boroughs of London*



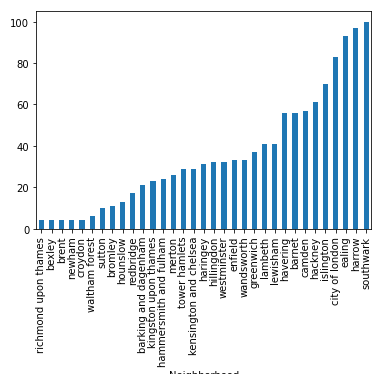
*Figure: All Boroughs of Inner London*

I used the Foursquare API to explore the boroughs and segment them. I designed the limit as 100 venue and the radius 500 meter for each borough from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude information from Foursquare API.



We can see that Southwark, Harrow and Ealing have reached the 100 limit of venues. On the other hand Richmond, Bexley, Brent, Newham, Croydon, Waltham Forest, Sutton are among the few boroughs that are below 20 venues in our given coordinates with Latitude and Longitude, in below graph.

The result doesn’t mean that inquiry run all the possible results in boroughs. Actually, it depends on given Latitude and Longitude information and here is we just run single Latitude and Longitude pair for each borough. We can increase the possibilities with Neighbourhood information with more Latitude and Longitude information



*Figure: Number of venues in each Borough*

In summary of this graph 222 unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for each borough in below table.

**

We have some common venue categories in boroughs. In this reason I used unsupervised learning **K-means** algorithm to cluster the boroughs. K-Means algorithm is one of the most common cluster method of unsupervised learning.

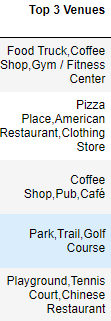
I will run K-Means to cluster the boroughs into 5.

**4. Results**

Here is my merged table with cluster labels for each borough.

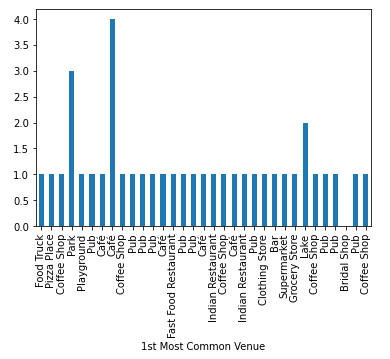


One of my aim was also show the number of top 3 venues information for each borough on the map. Thus, I grouped each borough by the number of top 3 venues and I combined those information in Join column.

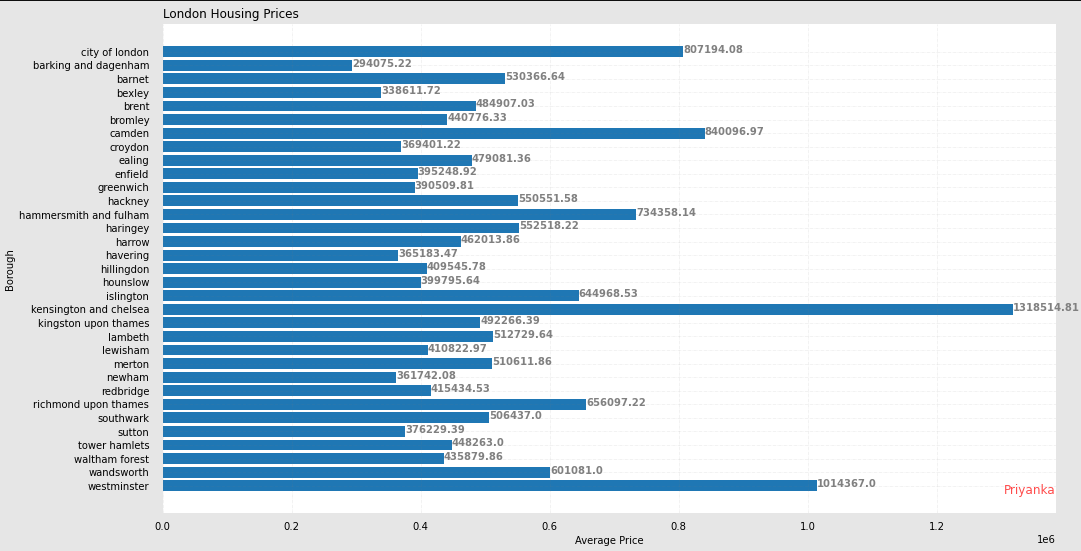


*Figure 1: Top 3 venues for each Borough in London.*

*Figure 2:* We can also estimate the number of 1st Most Common Venue in each cluster. Thus, we can create a bar chart which may help us to find proper labels for each cluster.



We can also examine that what is the frequency of average housing sales prices in different ranges. Thus, histogram can help to visualization:



**We can analyse Boroughs in London in three categories:**

**1. High Class Boroughs (~ 750,000+):**

-Kensington and Chelsea

-Westminster

-Camden

**2. Mid-Class Boroughs (300,000 – 750,000):**

-Barnet

-Islington

-Wands worth

**3. Lower Class Boroughs (200,000 – 300,000):**

-Barking and Dagenham

-Bexley

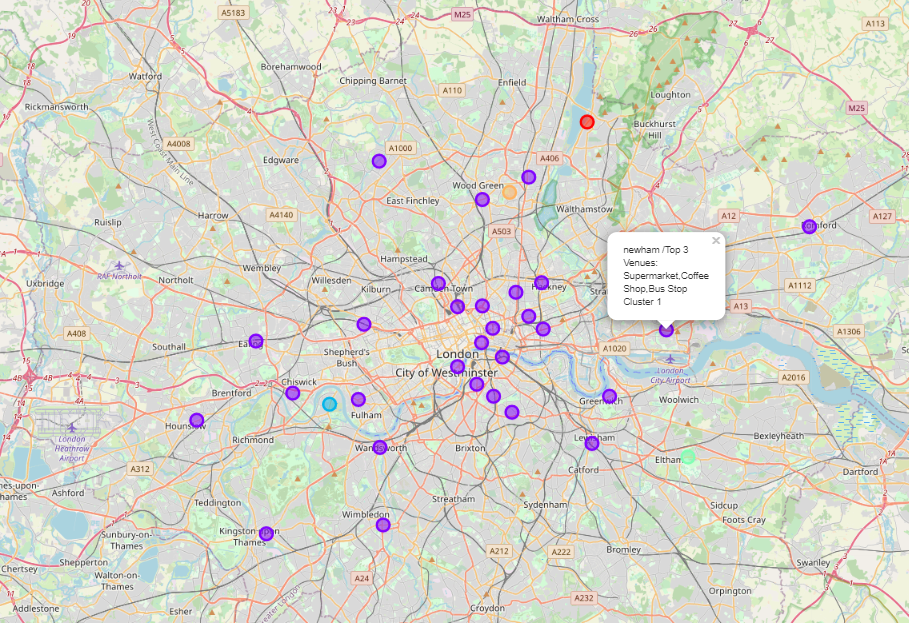
-Havering

On the basis of our budget one can easily choose among the Price-category of the Boroughs.

And on the basis of venue analysis we can come to a conclusion as to what category will benefit us the most in respective borough.

Let’s merge those new variables with related cluster information in our main master table.

You can now see join and see the last three ones in above table. You can also see a clustered map boroughs of London in the below.



**5. Discussion**

London is one of the largest metropolises in the world where over 9.3 million people live and it has a population density of 5,701 people per square kilometer. I’ve decided to use London in my project. The city is divided into local authority districts that make up the ceremonial county of Greater London; each is governed by a London borough council.

I used the Kmeans algorithm as part of this clustering study. I set the optimum k value to 5. For more detailed and accurate guidance, the data set can be expanded and the details of the neighborhood or street can also be drilled.

I also performed data analysis through this information by adding the coordinates of districts and home sales price averages as static data on GitHub. In future studies, these data can also be accessed dynamically from specific platforms or packages.

I ended the study by visualizing the data and clustering information on the London map. In future studies, web or telephone applications can be carried out to direct investors.

**6. Conclusion**

As a result, people are turning to big cities to start a business or work. For this reason, people can achieve better outcomes through their access to the platforms where such information is provided.

Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.

**7. References**

[1] London — Wikipedia

[2] London Boroughs House Price, Crime and Population data — Kaggle

[3] Foursquare API

[4] Google Map