

London Real Estate Analytics

Fuad Hassar

July 28, 2019

1. Introduction

1.1 Background

The city of London is the capital of UK and most populous city in UK. It provides lot of business opportunities and business friendly environment. It has attracted many people for living in London. It is a diverse and main tourist city in UK. It is a global hub of the world. The city is a major center for tourism, real estate, entertainment, theater, fashion, and the arts in the United Kingdom. This also means that the market of real estate business is highly competitive. As it is highly developed city, buying house is one of the highest cost. Thus, buying house in this fabulous city needs to be analyzed carefully. The insights derived from analysis will give good understanding of the proper cost of houses in London and will help us for clear strategic decision. Also, guide in reduction of risk and the return on investment to be reasonable.

1.2 Problem

In this scenario, it would be a great to adopt machine learning tools to assist homebuyers to make wise and effective decisions. As a result, the business problem that proposing is how could we provide support to homebuyers clientele to purchase a suitable real estate in London? in this uncertain economic and financial scenario?

To solve this business problem, we are going to cluster London neighborhoods in order to recommend venues and the current average price of real estate where homebuyers can make a real estate investment. We will recommend profitable venues according to amenities and essential facilities surrounding such venues, example elementary schools, high schools, hospitals & grocery stores.

1.3 Interest

Obviously, real estate agencies and investors who have interest in the investments in London or homebuyers would be very interested in accurate prediction of reasonable houses prices to help guide in reduction of risk and the return on investment.

2. Data acquisition and cleaning

2.1 Data sources

Data related to London properties including houses prices extracted from the Land Registry (<http://landregistry.data.gov.uk/>). The following fields comprise the address data included in Price paid Data: Postcode, Primary Addressable, Object Name. Street; Town; District. The Foursquare API will be used to explore locations across different venues according to the presence of amenities and essential facilities. The Foursquare explore function will be used to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. The k-means clustering algorithm will be used for the analysis. In the end, the Folium library will be used to visualize locations and facilities and their emerging clusters. In the end, recommendation of profitable real estate investments will be given.

2.2 Data cleaning

Data downloaded or scraped from mainly two sources were data got massaged and stored in different data frames. The data is very huge where decided to use data from 2017 until 2019. The Data columns got renamed and formatted to be suitable for machine learning algorithm according to the solution scope. Data got filtered based on city of London , list of street names in London, Average prices set to range between 2000000 to 25000000. In addition, Calculate the street-wise average price of the property and read the street-wise coordinates into a data frame. The data got Joined to find the coordinates of locations and according to clients budgets. Finally, visualization of map plot recommended locations on London map addressing all venues.

Table 1. Data Consolidation

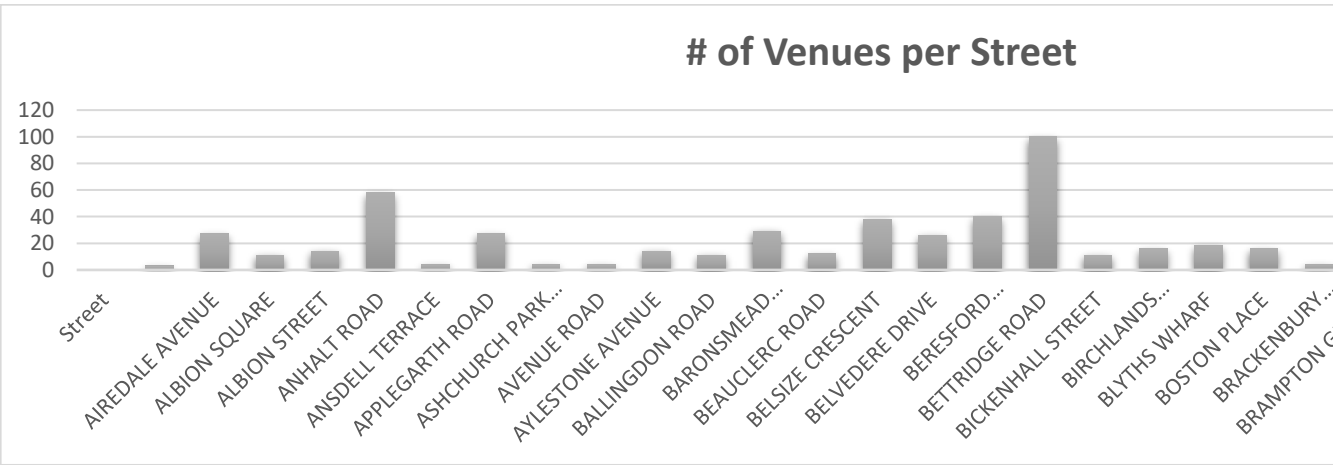
	Street	Avg_Price	city_coord
146	AIREDALE AVENUE	2.022500e+06	(53.8289048 , -1.8310423)
196	ALBION SQUARE	2.450000e+06	(-41.27375755, 173.289393239104)
197	ALBION STREET	2.096667e+06	(36.1659927 , -86.8074413)
391	ANHALT ROAD	2.435000e+06	(51.4803265 , -0.1667607)
406	ANSDELL TERRACE	2.250000e+06	(51.4998899 , -0.1891027)
421	APPLEGARTH ROAD	2.400000e+06	(53.749244 , -0.32678)
552	ASHCHURCH PARK VILLAS	2.150000e+06	(51.5000507 , -0.2421733)
671	AVENUE ROAD	2.143471e+06	(51.4067969 , -0.049519)
699	AYLESTONE AVENUE	2.286667e+06	(51.5409157 , -0.2178742)
760	BALLINGDON ROAD	2.105000e+06	(51.4541892 , -0.1588555)
853	BARONSMEAD ROAD	2.375000e+06	(51.4773147 , -0.239457)
979	BEAUCLERC ROAD	2.480000e+06	(51.4995771 , -0.2290331)
1089	BELSIZE CRESCENT	2.000000e+06	(51.5495675 , -0.1727325)
1100	BELVEDERE DRIVE	2.340000e+06	(51.4249173 , -0.2120774)
1132	BERESFORD TERRACE	2.100000e+06	(55.4571954 , -4.6291344)
1190	BETTRIDGE ROAD	2.025000e+06	(51.471074 , -0.2028354)
1213	BICKENHALL STREET	2.208500e+06	(51.5211969 , -0.1589341)
1251	BIRCHLANDS AVENUE	2.217000e+06	(51.4483941 , -0.1604676)
1380	BLYTHS WHARF	2.000000e+06	(51.50900105, -0.0345837265300945)
1431	BOSTON PLACE	2.167500e+06	(51.5073726, -0.340337807312508)
1511	BRACKENBURY GARDENS	2.150000e+06	(51.5006233, -0.230729)
1550	BRAMPTON GROVE	2.456875e+06	(51.5703648, -0.2833944)
1555	BRAMSHOT AVENUE	2.177900e+06	(51.4811602, 0.0226516)
1629	BRIARDALE GARDENS	2.397132e+06	(51.5601748, -0.1954305)

Exploratory Data Analysis

Started analyzing the outcome of consolidated data to understand and observe the number of venues nearby streets. We need to see the relation between number of venues compare to house prices. We have around 9289 records associated to multiple venues.

Table 2. Venues Counts per Street

	Street Latitude	Street Longitude	Venue	Venue Latitude	Venue Longitude
Street					
AIREDALE AVENUE	3	3	3	3	3
ALBION SQUARE	27	27	27	27	27
ALBION STREET	11	11	11	11	11
ANHALT ROAD	14	14	14	14	14
ANSDELL TERRACE	58	58	58	58	58
APPLEGARTH ROAD	4	4	4	4	4
ASHCHURCH PARK VILLAS	27	27	27	27	27
AVENUE ROAD	4	4	4	4	4
AYLESTONE AVENUE	4	4	4	4	4
BALLINGDON ROAD	14	14	14	14	14
BARONSMEAD ROAD	11	11	11	11	11
BEAUCLERC ROAD	29	29	29	29	29
BELSIZE CRESCENT	12	12	12	12	12
BELVEDERE DRIVE	38	38	38	38	38
BERESFORD TERRACE	26	26	26	26	26
BETTRIDGE ROAD	40	40	40	40	40
BICKENHALL STREET	100	100	100	100	100
BIRCHLANDS AVENUE	11	11	11	11	11
BLYTHS WHARF	16	16	16	16	16
BOSTON PLACE	18	18	18	18	18
BRACKENBURY GARDENS	16	16	16	16	16
BRAMPTON GROVE	4	4	4	4	4
BRAMSHOT AVENUE	7	7	7	7	7
BRIARDALE GARDENS	7	7	7	7	7
BROADLANDS ROAD	2	2	2	2	2



Around 252 potential streets allocated. For better insight, we have calculated the average of venues per street.

Table 3. Average Venue

	Street	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Amphit
0	AIREDALE AVENUE	0.000000	0.0	0.0	0.000000	0.000000	0.000000
1	ALBION SQUARE	0.000000	0.0	0.0	0.000000	0.000000	0.000000
2	ALBION STREET	0.000000	0.0	0.0	0.000000	0.090909	0.090909
3	ANHALT ROAD	0.000000	0.0	0.0	0.000000	0.000000	0.000000
4	ANSDELL TERRACE	0.000000	0.0	0.0	0.000000	0.000000	0.000000
5	APPLEGARTH ROAD	0.000000	0.0	0.0	0.000000	0.000000	0.000000
6	ASHCHURCH PARK VILLAS	0.000000	0.0	0.0	0.000000	0.000000	0.000000
7	AVENUE ROAD	0.000000	0.0	0.0	0.000000	0.000000	0.000000
8	AYLESTONE AVENUE	0.000000	0.0	0.0	0.000000	0.000000	0.000000

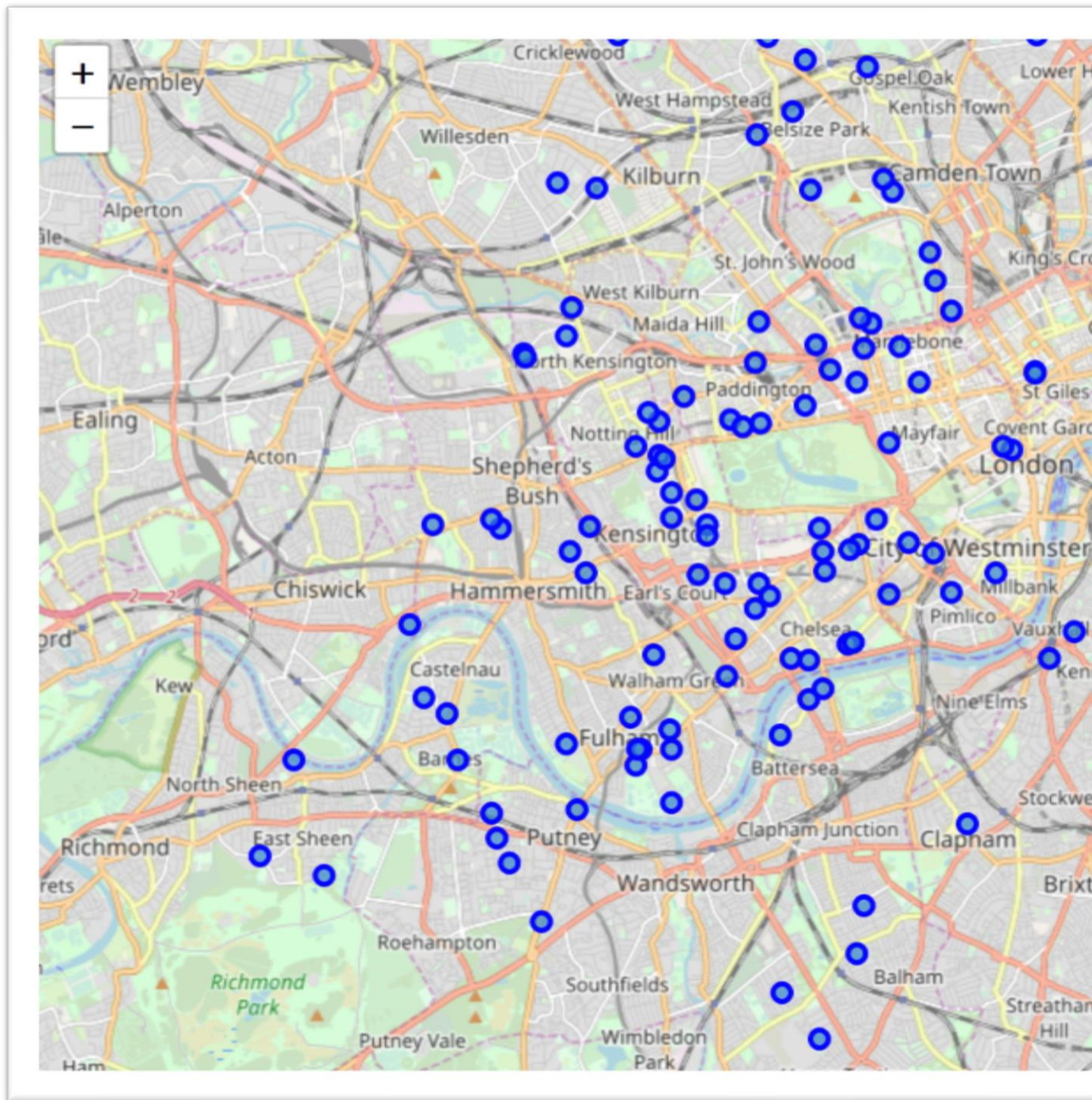
The most common top venues/facilities nearby profitable real estate investments are allocated for better visibility of homebuyer's interest.

Table 4. Top Venues

	Street	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0	AIREDALE AVENUE	Bakery	Harbor / Marina	Business Service	Zoo Exhibit	Food & Drink Shop	Fast Food Restaurant	Filipino Restaurant
1	ALBION SQUARE	Café	Restaurant	Bar	Indian Restaurant	Coffee Shop	Pub	French Restaurant
2	ALBION STREET	BBQ Joint	Pizza Place	Coffee Shop	Auto Workshop	Athletics & Sports	Lounge	Smoothie Shop
3	ANHALT ROAD	Pub	Grocery Store	Japanese Restaurant	Garden	Gym / Fitness Center	English Restaurant	Diner
4	ANSDELL TERRACE	Clothing Store	Italian Restaurant	Restaurant	Hotel	Juice Bar	Pub	Chinese Restaurant

Map visulaization is set to see the distribution of locations with respect to venues

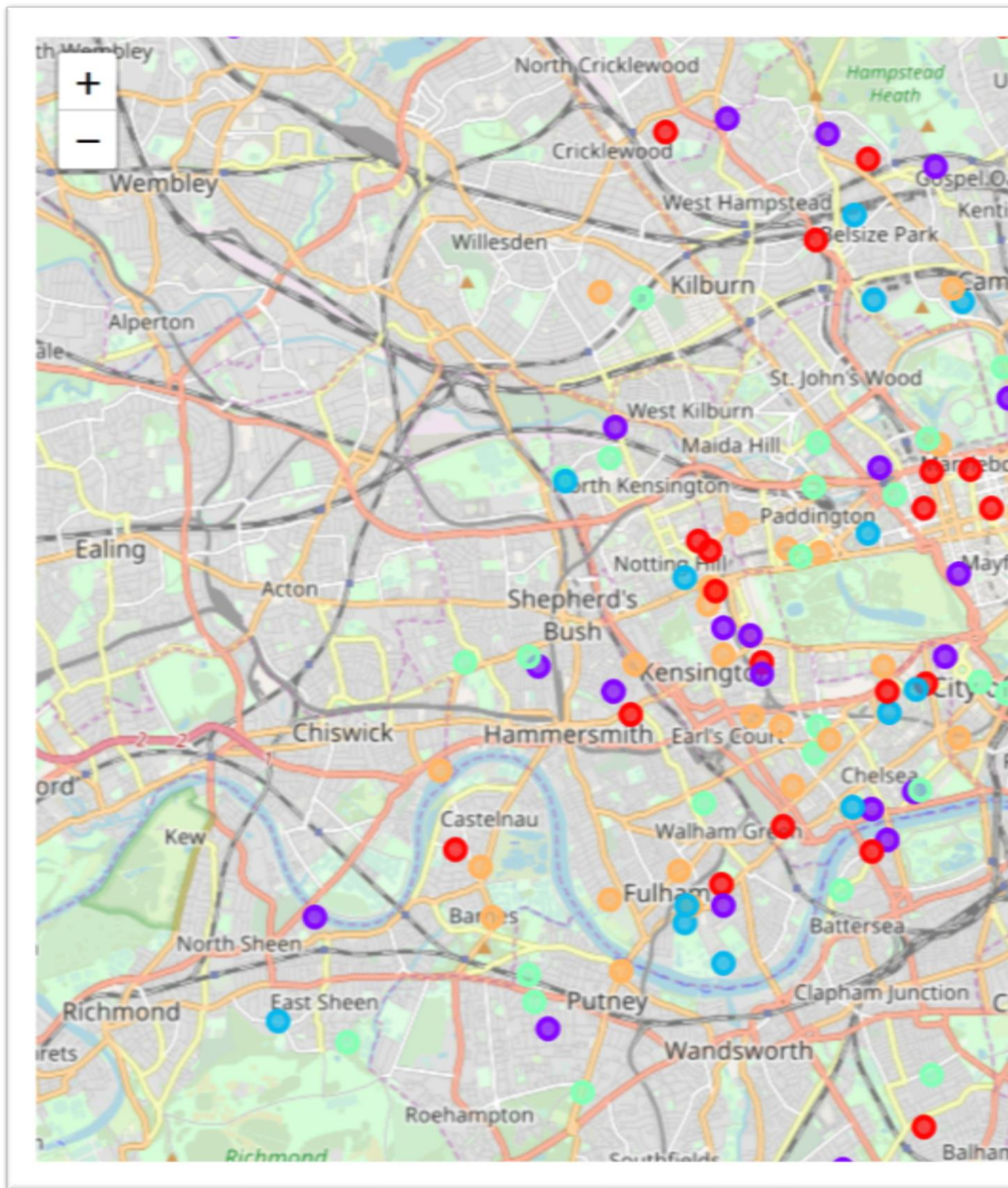
Map 1. Street/ Venues



4. Predictive Modeling

The K-Mean Clustering model is being used for London Real Estate project. We have applied clustering method to help recommend best investment and recommendation according to surrounding facilities. k-means clustering technique will be used to analyze real estates data for London. Before, venues around areas will be identified through foursquare API.

Map 2. Clusters



With a population of more than 8.6 million, London is a densely populated metropolis with a melting pot of multi-ethnic residents from all over the world. As the hub for the UK's economy, politics and culture this city attracts a great deal of Brits and foreigners, despite high costs of living and higher than average housing prices than the rest of the UK.

Buying UK property is likely to be one of the biggest investments of your life, particularly in London where house prices are considerably higher. And with so many people all looking for accommodation in London, the housing market is very competitive for buying London property, with the market operating under its own influences and factors.

prices vary depending on where and the type of London property you buy.

In our analysis, outcome divided into 5 clustered included prices and common venues., Each cluster shows the most common venues targeted as example Clusters 4, indicated that target home buyers live in 'green' areas with parks, theaters while cluster 0, target restaurants and cafes. The areas within Central London are often a preferred choice a close proximity to everything.

Cluster No 0:

	Avg_Price	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
355	2435000.0	Pub	Plaza	Pizza Place	Grocery Store
381	2400000.0	Bar	Pub	Casino	Nightclub
1439	2397132.0	Food Truck	Indian Restaurant	Grocery Store	Gym / Fitness Center
1894	2425000.0	Hotel	Indian Restaurant	French Restaurant	Chinese Restaurant
2171	2414000.0	Coffee Shop	Bus Station	Zoo	English Restaurant

Cluster No 1:

	Avg_Price	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
368	2250000.0	Clothing Store	Italian Restaurant	Café	English Restaurant
1094	2200000.0	Pub	Soccer Stadium	Italian Restaurant	Grocery Store
1108	2217000.0	Pub	French Restaurant	Train Station	Bar
1950	2200000.0	Nightlife	Soccer Field	Fish & Chips	English Restaurant

Cluster No 2:

	Avg_Price	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
20	2.367093e+06	Grocery Store	Park	Waterfront	Hotel
753	2.375000e+06	Food & Drink Shop	Restaurant	Breakfast Spot	Nightlife Spot
1079	2.351667e+06	Hotel	Chinese Restaurant	Gastropub	Restaurant
1820	2.375000e+06	Pub	Bakery	Park	Hotel
1871	2.352889e+06	Pub	Bakery	Pizza Place	Hotel

Cluster No 3:

	Avg_Price	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
368	2250000.0	Clothing Store	Italian Restaurant	Café	Entertainment Venue
1094	2200000.0	Pub	Soccer Stadium	Italian Restaurant	Grocery Store
1108	2217000.0	Pub	French Restaurant	Train Station	Bar
1950	2200000.0	Nightlife Spot	Soccer Field	Fish & Chips Shop	Entertainment Venue
2306	2250000.0	Gastropub	Construction & Landscaping	Health & Beauty Service	Fish & Chips Shop

Cluster No 4:

	Avg_Price	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
178	2.450000e+06	Café	Restaurant	Indian Restaurant	Bar
867	2.480000e+06	Coffee Shop	Pub	Hotel	Tea Room
1371	2.475833e+06	Bar	Lake	Middle Eastern Restaurant	Museum
2237	2.495000e+06	NaN	NaN	NaN	NaN
2278	2.450000e+06	Garden	Cocktail Bar	Gym / Fitness Center	Park

5. Conclusions

We have gathered data related to London properties including prices paid from Land Registry website. The places got explored across different locations in London and according to different venues including amenities and facilities. Data has been extracted from FourSquare APIs and got sorted out and arranged for visualization. Accordingly, we were able to recommend proper profitable real estate investment.

As a result, It has been found that areas like Notting Hill, Kensington, Marylebone, Brompton considered highly profitable venues to purchase a real estate according to amenities and essential facilities surrounding such venues i.e. Parks, supermarket, schools, hospitals, etc. In the other hand, Chelsea, Wandsworth, Balham and Fulham are next future elite venues with a wide range of amenities and facilities.

The result was divided into 5 clustered and it has been analyzed, each cluster shows the most common venues targeted as example Clusters 4, indicated that target home buyers live in 'green' areas with parks, theaters while cluster 0, target restaurants and cafes.

