THE BATTLE OF NEIGHBORHOODS

CLUSTER ANALYSIS OF REAL ESTATE IN LONDON

BUYING REAL ESTATE IN LONDON IN 2018

TUAMSEN DEAN DAJAU

MAY 2020

1. INTRODUCTION: BUSINESS PROBLEM

1.1 Background

According to Bloomberg News in 2018, the London Housing Market is in a rut. It was facing a number of different headwinds, including the prospect of higher taxes and a warning from the Bank of England that U.K. home values could fall as much as 30 per cent in the event of a disorderly exit from the European Union. More specifically, four overlooked cracks suggest that the London market may be in worse shape than many realize: hidden price falls, record-low sales, homebuilder exodus and tax hikes addressing overseas buyers of homes in England and Wales.

1.2 Business Problem

In this scenario, it is urgent to adopt machine learning tools in order to assist homebuyers' clientele in London to make wise and effective decisions. As a result, the business problem we are currently posing is: how could we have provided support to homebuyers' clientele to purchase a suitable real estate in London during this uncertain economic and financial scenario?

To solve this business problem, I am 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. I will recommend profitable venues according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores.

2. THE DATA

Data on London properties and the relative price paid data were extracted from the HM Land Registry (http://landregistry.data.gov.uk/). The following fields comprise the address data included in Price Paid Data: Postcode; PAON Primary Addressable Object Name. Typically the house number or name; SAON, Secondary Addressable Object Name. If there is a sub-building, for example, the building is divided into flats, there will be a SAON; Street; Locality; Town/City; District; County.

To explore and target recommended locations across different venues according to the presence of amenities and essential facilities, we will access data through FourSquare

API interface and arrange them as a dataframe for visualization. By merging data on London properties and the relative price paid data from the HM Land Registry and data on amenities and essential facilities surrounding such properties from FourSquare API interface, we will be able to recommend profitable real estate investments.
The figures below represent data from the HM Land Registry and data on amenities and essential facilities from FourSquare API interface:

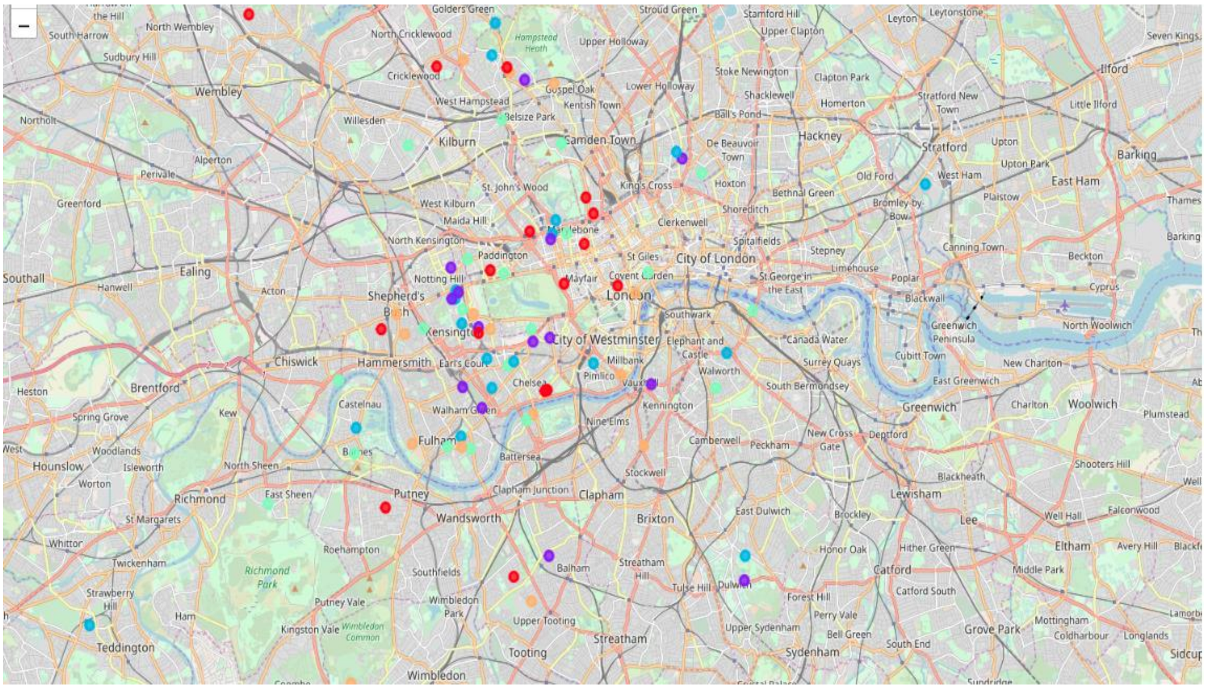
Head of London's data frame before data cleaning:

	{79A74E21-D11E-1289-E053-6B04A8C01627}	770000	2018-09-25 00:00	SK7 1AR	D	N	F	5	Unnamed: 8	OAK MEADOW	BRAMHALL	STOCKPORT	STOCKPORT.1	GREATER MANCHESTER	A	A.1
0	{79A74E21-D11F-1289-E053-6B04A8C01627}	253500	2018-09-24 00:00	M6 8GQ	D	N	F	1	NaN	RIVINGTON ROAD	NaN	SALFORD	SALFORD	GREATER MANCHESTER	A	A
1	{79A74E21-D120-1289-E053-6B04A8C01627}	231950	2018-09-28 00:00	WA3 2UE	D	Y	F	35	NaN	STONEACRE CLOSE	LOWTON	WARRINGTON	WIGAN	GREATER MANCHESTER	A	A
2	{79A74E21-D121-1289-E053-6B04A8C01627}	112500	2018-08-29 00:00	OL6 6RJ	S	N	F	102	NaN	THORNFIELD GROVE	NaN	ASHTON-UNDER-LYNE	TAMESIDE	GREATER MANCHESTER	A	A
3	{79A74E21-D122-1289-E053-6B04A8C01627}	184995	2018-06-15 00:00	M46 0TW	S	Y	F	37	NaN	THREADNEEDLE PLACE	ATHERTON	MANCHESTER	WIGAN	GREATER MANCHESTER	A	A
4	{79A74E21-D123-1289-E053-6B04A8C01627}	214995	2018-09-28 00:00	M28 3XS	D	Y	L	9	NaN	MARPLE GARDENS	WORSLEY	MANCHESTER	SALFORD	GREATER MANCHESTER	A	A

Head of data frame showing London's affordable streets with average price and location:

	Street	Avg_Price	city_coord	Latitude	Longitude
196	ALBION SQUARE	2.450000e+06	(-41.27375755, 173.28939323910353)	-41.273758	173.289393
390	ANHALT ROAD	2.435000e+06	(51.4803164, -0.1668011)	51.480316	-0.166801
405	ANSDELL TERRACE	2.250000e+06	(51.4998899, -0.1891027)	51.499890	-0.189103
422	APPLEGARTH ROAD	2.400000e+06	(53.7486539, -0.3266704)	53.748654	-0.326670
855	BARONSMEAD ROAD	2.375000e+06	(51.4773147, -0.239457)	51.477315	-0.239457
981	BEAUCLERC ROAD	2.480000e+06	(30.2114523, -81.6179807)	30.211452	-81.617981
1102	BELVEDERE DRIVE	2.340000e+06	(38.0728178, -78.4587964)	38.072818	-78.458796
1215	BICKENHALL STREET	2.208500e+06	(51.5212014, -0.1589082)	51.521201	-0.158908
1253	BIRCHLANDS AVENUE	2.217000e+06	(51.4483941, -0.1604676)	51.448394	-0.160468
1553	BRAMPTON GROVE	2.456875e+06	(51.5899607, -0.3185249)	51.589961	-0.318525
1632	BRIARDALE GARDENS	2.397132e+06	(51.5601748, -0.1954305)	51.560175	-0.195431
1797	BROOKWAY	2.400000e+06	(45.432184899999996, -122.80281166115779)	45.432185	-122.802812
1914	BURBAGE ROAD	2.445000e+06	(52.5385071, -1.3536736)	52.538507	-1.353674
1980	BURY WALK	2.492500e+06	(52.1455294, -0.4235933)	52.145529	-0.423593
2068	CALLCOTT STREET	2.375000e+06	(51.5083499, -0.1983276)	51.508350	-0.198328
2129	CAMPDEN HILL ROAD	2.379653e+06	(51.5081106, -0.1996673)	51.508111	-0.199667
2136	CAMPION ROAD	2.461000e+06	(52.681078, 0.9655987)	52.681078	0.965599
2158	CANNING PLACE	2.425000e+06	(52.6410267, -1.135223)	52.641027	-1.135223

Map visualization of the data frame:



3. **METHODOLOGY**

The Methodology section will describe the main components of our analysis and predication system. The Methodology section comprises four stages:

- Collect Inspection Data
- Explore and Understand Data
- Data preparation and preprocessing
- Modeling

4. **RESULTS**

I will discuss the results under two main perspectives.

First, I examined them according to neighborhoods/London areas. It is interesting to note that, although West London (Notting Hill, Kensington, Chelsea, Marylebone) and North-West London (Hampsted) might be considered highly profitable venues to purchase a real estate according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores, South-West London (Wandsworth, Balham) and North-West London (Isliington) are arising as next future elite venues with a wide range of amenities and facilities. Accordingly, one might target under-priced real estates in these areas of London in order to make a business affair.

Second, I analyzed the results according to the five clusters we have produced. Even though, all clusters could praise an optimal range of facilities and amenities, we have found two main patterns. The first pattern we are referring to, i.e. Clusters 0, 2 and 4, may target home buyers prone to live in 'green' areas with parks, waterfronts. Instead, the second pattern we are referring to, i.e. Clusters 1 and 3, may target individuals who love pubs, theatres and soccer.

Cluster 0:

	Avg_Price	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
405	2250000.0	Restaurant	Clothing Store	Hotel	Juice Bar	Pub	Italian Restaurant	Bakery	Garden	Indian Restaurant	Recording Studio
1215	2208500.0	Café	Restaurant	Pizza Place	Hotel	Garden	Gastropub	Movie Theater	Italian Restaurant	Bakery	Bar
1253	2217000.0	Pub	French Restaurant	Breakfast Spot	Coffee Shop	Chinese Restaurant	Lake	Train Station	Bakery	Brewery	Fish Market
2225	2200000.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2638	2250000.0	Bakery	Coffee Shop	Theater	Supermarket	Grocery Store	Gym / Fitness Center	Pharmacy	Restaurant	Pizza Place	Bookstore

Cluster 2:

	Avg_Price	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
855	2375000.0	Indie Movie Theater	Thai Restaurant	Community Center	Pub	Coffee Shop	Restaurant	Park	Café	Farmers Market	Nature Preserve
1102	2340000.0	Playground	Pool	Athletics & Sports	Zoo	Farm	Electronics Store	English Restaurant	Ethiopian Restaurant	Event Space	Exhibit
2068	2375000.0	Pub	Park	Pizza Place	Hotel	Grocery Store	Ice Cream Shop	Indian Restaurant	Yoga Studio	Mexican Restaurant	Tennis Court
2129	2379652.7	Pub	Grocery Store	Park	Hotel	Coffee Shop	Pizza Place	Bakery	Ice Cream Shop	Indian Restaurant	Yoga Studio
2944	2367500.0	Hotel	Pub	Garden	Café	Coffee Shop	Italian Restaurant	Bar	Mediterranean Restaurant	Chinese Restaurant	Tapas Restaurant

Cluster 4:

	Avg_Price	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2242	2.300000e+06	Soup Place	Farm	Zoo	Farmers Market	Electronics Store	English Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Factory
2406	2.286679e+06	Café	Pub	Coffee Shop	Italian Restaurant	Bar	Convenience Store	French Restaurant	Park	Market	Cafeteria
2686	2.287500e+06	Pub	Gym / Fitness Center	Brewery	Gift Shop	Art Museum	Zoo	Farm	English Restaurant	Ethiopian Restaurant	Event Space
3377	2.298000e+06	Hotel	Zoo	Farmers Market	Electronics Store	English Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Factory	Falafel Restaurant
4285	2.265000e+06	Pub	Farm	Egyptian Restaurant	Electronics Store	English Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Factory	Falafel Restaurant

Cluster 1:

	Avg_Price	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
196	2450000.0	Café	Coffee Shop	Bar	Indian Restaurant	Restaurant	Pub	Brewery	Supermarket	Beer Garden	Seafood Restaurant
981	2480000.0	Spa	Automotive Shop	Pizza Place	Farmers Market	English Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Factory	Falafel Restaurant
1553	2456875.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1980	2492500.0	Supermarket	English Restaurant	Dry Cleaner	Hardware Store	Park	American Restaurant	Pub	Fast Food Restaurant	Gym	Coffee Shop
2136	2461000.0	Windmill	Soccer Field	Bus Station	Zoo	Farmers Market	English Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Factory

Cluster 3:

	Avg_Price	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
390	2435000.0	Pub	Grocery Store	French Restaurant	Art Gallery	Japanese Restaurant	Diner	Cocktail Bar	Gym / Fitness Center	Garden	Italian Restaurant
422	2400000.0	Casino	Bar	Auto Dealership	Nightclub	Farmers Market	Ethiopian Restaurant	Event Space	Exhibit	Factory	Falafel Restaurant
1632	2397132.0	Gym / Fitness Center	Breakfast Spot	Coffee Shop	Grocery Store	Zoo	Farm	Electronics Store	English Restaurant	Ethiopian Restaurant	Event Space
1797	2400000.0	Home Service	Zoo	Farmers Market	Electronics Store	English Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Factory	Falafel Restaurant
1914	2445000.0	Grocery Store	Bar	Athletics & Sports	Dance Studio	Farmers Market	English Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Factory

5. CONCLUSION

To sum up, according to Bloomberg News in 2018, the London Housing Market was in a rut. It was facing a number of different headwinds, including the prospect of higher taxes and a warning from the Bank of England that U.K. home values could fall as much as 30 percent in the event of a disorderly exit from the European Union. In this scenario, it is urgent to adopt machine learning tools in order to assist homebuyers' clientele in London to make wise and effective decisions.

To solve this business problem, I clustered London neighborhoods in order to recommend venues and the current average price of real estate where homebuyers can make a real estate investment. We recommended profitable venues according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores.

First, I gathered data on London properties and the relative price paid data were extracted from the HM Land Registry (http://landregistry.data.gov.uk/). Moreover, to explore and target recommended locations across different venues according to the presence of amenities and essential facilities, I accessed data through FourSquare API interface and arranged them as a data frame for visualization. By merging data on London properties and the relative price paid data from the HM Land Registry and data on amenities and essential facilities surrounding such properties from FourSquare API interface, I was able to recommend profitable real estate investments.

Second, The Methodology section comprised four stages: 1. Collect Inspection Data; 2. Explore and Understand Data; 3. Data preparation and preprocessing; 4. Modeling. Particularly in the modeling section, I used the k-means clustering technique as it is fast and efficient in terms of computational cost, is highly flexible to account for mutations in real estate market in London and is accurate.

Finally, I drew the conclusion that even though the London Housing Market may be in a rut, it is still an "ever-green" for business affairs. I discussed the results under two main perspectives. First, I examined them according to neighborhoods/London areas. Now although West London (Notting Hill, Kensington, Chelsea, Marylebone) and North-West London (Hampsted) might be considered highly profitable venues to purchase a real estate according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores, South-West London (Wandsworth, Balham) and North-West London (Isliington) are arising as next future elite venues with a wide range of amenities and facilities. Accordingly, one might target under-priced real estates in these areas of London in order to make a business affair. Second, I analyzed the results according to the five clusters we produced. While Clusters 0, 2 and 4 may target home buyers prone to live in 'green' areas with parks, waterfronts, Clusters 1 and 3 may target individuals who love pubs, theatres and soccer.