Capstone Project - The Battle of Neighborhoods

Introduction

This is to study and come up with profitable real estate housing options in London neighborhoods.

Background

According to Bloomberg News, the London Housing Market is in a rut. It is now 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. 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.

Business Problem

In this scenario, it is urgent to adopt machine learning tools in order to assist home buyer clients in London to make wise and effective decisions. As a result, the business problem we are currently posting is: how could we provide support to home buyers clients in order 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 real estate prices where homebuyers can make a real estate investment. We will recommend profitable venues according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores.

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 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, we will be able to recommend profitable real estate investments.

Methodology

Below are the main components of our analysis and predication system.

1. Collect & Inspect Data
2. Explore and Understand Data
3. Data preparation and preprocessing
4. Modeling

1. Collect & Inspect Data

Now let's download and explore the data. Get the data for examination from the Source: http://landregistry.data.gov.uk/

Out[2]:

	{666758D7-43A9-3363-E053-6B04A8C0D74E}	405000	2018-01-25 00:00	WR15 8LH	D	N	F	RAMBLERS WAY	Unnamed: 8	Unnamed: 9	BORASTON	TENBURY WELLS	SHROPSHIRE	SHROPSHIRI
0	{666758D7-43AA-3363-E053-6B04A8C0D74E}	315000	2018-01-23 00:00	SY7 8QA	D	N	F	MONT CENISE	NaN	NaN	CLUN	CRAVEN ARMS	SHROPSHIRE	SHROPSHIRI
1	{666758D7-43AD-3363-E053-6B04A8C0D74E}	165000	2018-01-19 00:00	SY1 2BF	T	Y	F	42	NaN	PENSON WAY	NaN	SHREWSBURY	SHROPSHIRE	SHROPSHIRI
2	{666758D7-43B0-3363-E053-6B04A8C0D74E}	370000	2018-01-22 00:00	SY8 4DF	D	N	F	WILLOW HEY	NaN	NaN	ASHFORD CARBONEL	LUDLOW	SHROPSHIRE	SHROPSHIRI
3	{666758D7-43B3-3363-E053-6B04A8C0D74E}	320000	2018-01-19 00:00	TF10 7ET	D	N	F	3	NaN	PRINCESS GARDENS	NaN	NEWPORT	WREKIN	WREKIN

2. Explore and Understand Data

To explore and target recommended locations across different venues according to the presence of amenities and essential facilities, we will access data through Four Square API interface and arrange 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 Four Square API interface, we will be able to recommend profitable real estate investments.

Our original dataset consists of over 1020000 rows and 16 features

3. Data preparation and preprocessing

At this stage, we prepare our dataset for the modeling process, opt for the most suitable machine learning algorithm for our scope.

Accordingly, we performed the following steps:

1. Rename the column names
2. Format the date column
3. Sort data by date of sale
4. Select data only for the city of London
5. Make a list of street names in London
6. Calculate the street-wise average price of the property
7. Read the street-wise coordinates into a data frame, eliminating recurring word London from individual names
8. Join the data to find the coordinates of locations which fit into client's budget
9. Plot recommended locations on London map along with current market prices.

Out[6]:

	Street	Avg_Price	city_coord
196	ALBION SQUARE	2.450000e+06	(-41.27375755, 173.289393239104)
391	ANHALT ROAD	2.435000e+06	(51.4803265, -0.1667607)
406	ANSDELL TERRACE	2.250000e+06	(51.4998899, -0.1891027)
422	APPLEGARTH ROAD	2.400000e+06	(53.7486539, -0.3266704)
855	BARONSMEAD ROAD	2.375000e+06	(51.4773147, -0.239457)
981	BEAUCLERC ROAD	2.480000e+06	(51.4995771, -0.2290331)
1102	BELVEDERE DRIVE	2.340000e+06	(44.7628418, -63.6692314)
1215	BICKENHALL STREET	2.208500e+06	(51.5211969, -0.1589341)
1253	BIRCHLANDS AVENUE	2.217000e+06	(51.4483941, -0.1604676)

We can now proceed to the Modeling phase. We will analyze neighborhoods to recommend real estates where home buyers can make a real estate investment. We will then recommend profitable venues according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores.

4.Modeling

After exploring the dataset and gaining insights into it, we are ready to use the clustering methodology to analyze real estates. We will use 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.

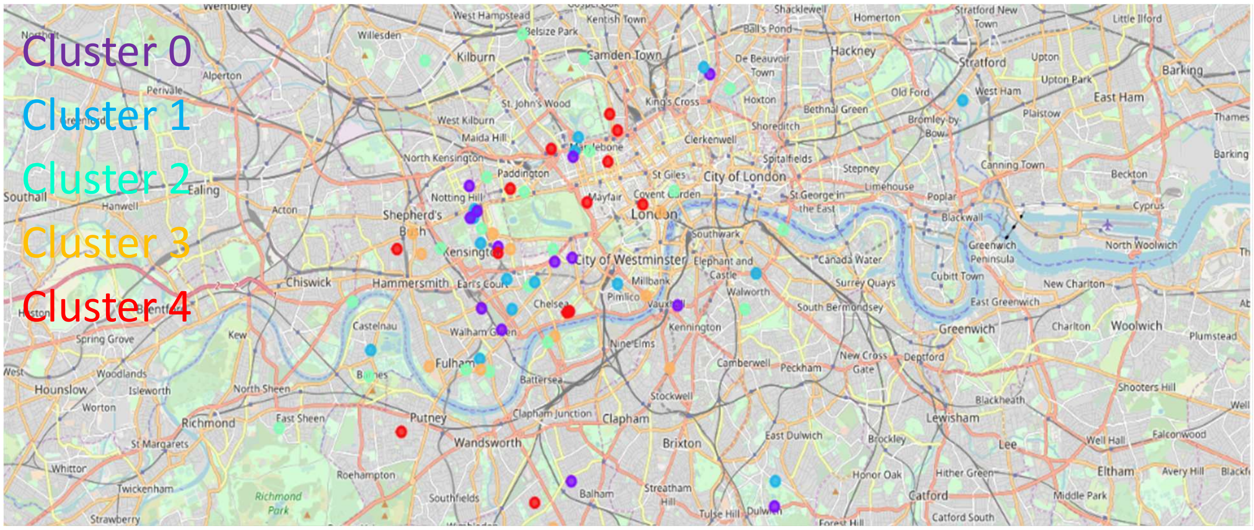
A glimpse of the data set with the most common venues/facilities nearby real estate investments.

Out[30]:

	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	ALBION SQUARE	Café	Pub	Bar	Coffee Shop	Indian Restaurant	Restaurant	Burger Joint	Supermarket	Seafood Restaurant	Brewery
1	ANHALT ROAD	Pub	Grocery Store	Japanese Restaurant	Garden	Gym / Fitness Center	Diner	Plaza	English Restaurant	Pizza Place	Cocktail Bar
2	ANSDELL TERRACE	Clothing Store	Restaurant	Italian Restaurant	Café	Hotel	Pub	Chinese Restaurant	Indian Restaurant	French Restaurant	Garden
3	APPLEGARTH ROAD	Pub	Nightclub	Casino	Food	Farmers Market	Fast Food Restaurant	Filipino Restaurant	Fish & Chips Shop	Fish Market	Flea Market
4	AYLESTONE AVENUE	Park	Café	Movie Theater	Yoga Studio	Food	Farmers Market	Fast Food Restaurant	Filipino Restaurant	Fish & Chips Shop	Fish Market

152 venues were analyzed based on 350 categories of facilities.

After our inspection of venues/facilities/amenities nearby the most profitable real estate investments in London, we performed clustering properties by venues/facilities/amenities nearby into 5 clusters using K-Means.



Results and Discussion

First of all, even though the London Housing Market may be in a rut, it is still an "ever-green" for business affairs.

We may discuss our results under two main perspectives.

First, we may examine 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 underpriced real estates in these areas of London in order to make a business affair.

Second, we may analyze our 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.

Conclusion

To sum up, according to Bloomberg News, the London Housing Market is in a rut. It is now 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 home

buyer clients in London to make wise and effective decisions. As a result, the business problem we were posing was: how could we provide support to homebuyer clients in to purchase a suitable real estate in London in this uncertain economic and financial scenario?

To solve this business problem, we 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, we 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, we 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, we were 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. In particular, in the modeling section, we 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, we drew the conclusion that even though the London Housing Market may be in a rut, it is still an "ever-green" for business affairs. We discussed our results under two main perspectives. First, we examined them according to neighborhoods/London areas. 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, we analyzed our 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.