London Property Analysis

M. Faizan Tahir

Introduction

Description of the Problem

In this scenario, it is urgent to adopt machine learning tools in order to assist home buyers clientele in London to make wise and effective decisions. As a result, the business problem we are currently posing is: how could we provide support to home buyers' clientele in 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 i.e. elementary schools, high schools, hospitals & grocery stores.

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.

Target Audience

Considering the diversity of London, there is a high multicultural sense. London is a place where different shades live. As such, in the search for an high-end African-inclined restaurant, there is a high shortage. The target audience is broad, it ranges from Londoners, tourists and those who are passionate about organic food.

Data Section

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.

Methodology

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

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

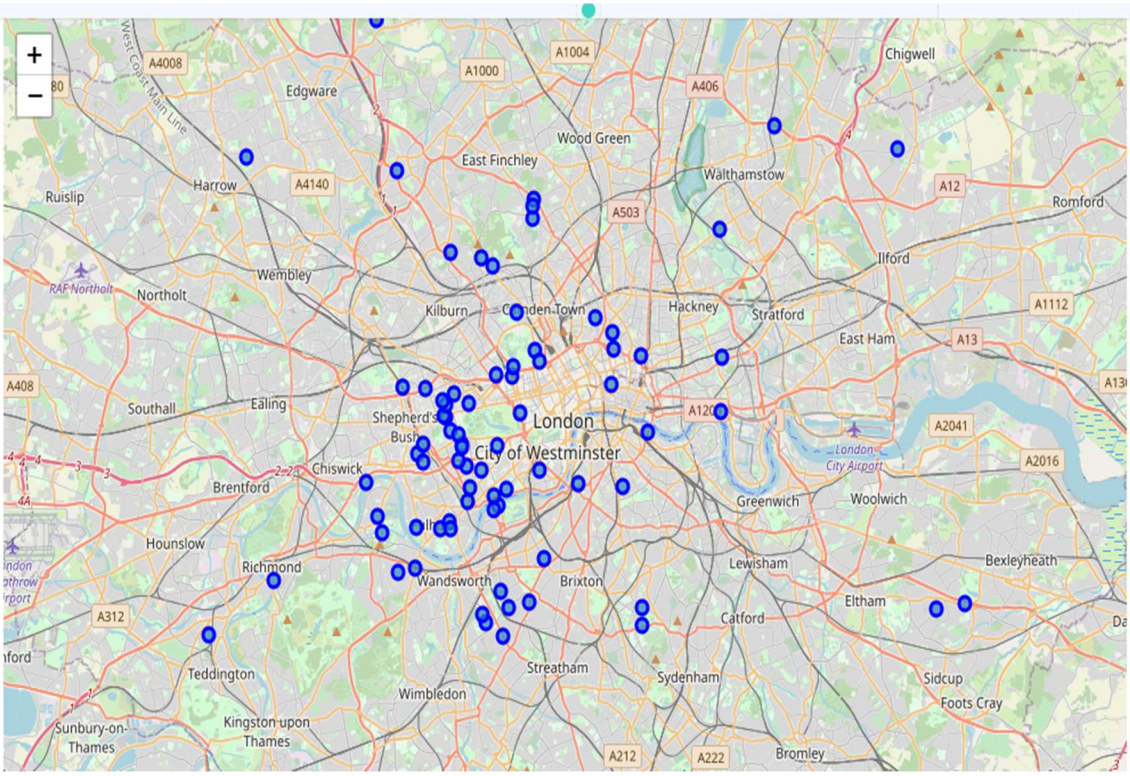
Explore and Understand Data

We read the dataset that we collected from the HM Land Registry website into a pandas' data frame and display the first five rows

	Street	Avg_Price	Latitude	Longitude
196	ALBION SQUARE	2.450000e+06	-41.273758	173.289393
390	ANHALT ROAD	2.435000e+06	51.480316	-0.166801
405	ANSDELL TERRACE	2.250000e+06	51.499890	-0.189103
422	APPLEGARTH ROAD	2.400000e+06	53.748654	-0.326670
855	BARONSMEAD ROAD	2.375000e+06	51.477315	-0.239457
981	BEAUCLERC ROAD	2.480000e+06	30.211452	-81.617981
1102	BELVEDERE DRIVE	2.340000e+06	38.072818	-78.458796
1215	BICKENHALL STREET	2.208500e+06	51.521201	-0.158908
1253	BIRCHLANDS AVENUE	2.217000e+06	51.448394	-0.160468
1553	BRAMPTON GROVE	2.456875e+06	51.589961	-0.318525
1632	BRIARDALE GARDENS	2.397132e+06	51.560175	-0.195431
1797	BROOKWAY	2.400000e+06	45.432185	-122.802812
1914	BURBAGE ROAD	2.445000e+06	52.538507	-1.353674
1980	BURY WALK	2.492500e+06	52.145529	-0.423593
2068	CALLCOTT STREET	2.375000e+06	51.508350	-0.198328

We can now proceed to the Modeling phase. We will analyze neighborhoods to recommend real estate 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.

Out[24]:



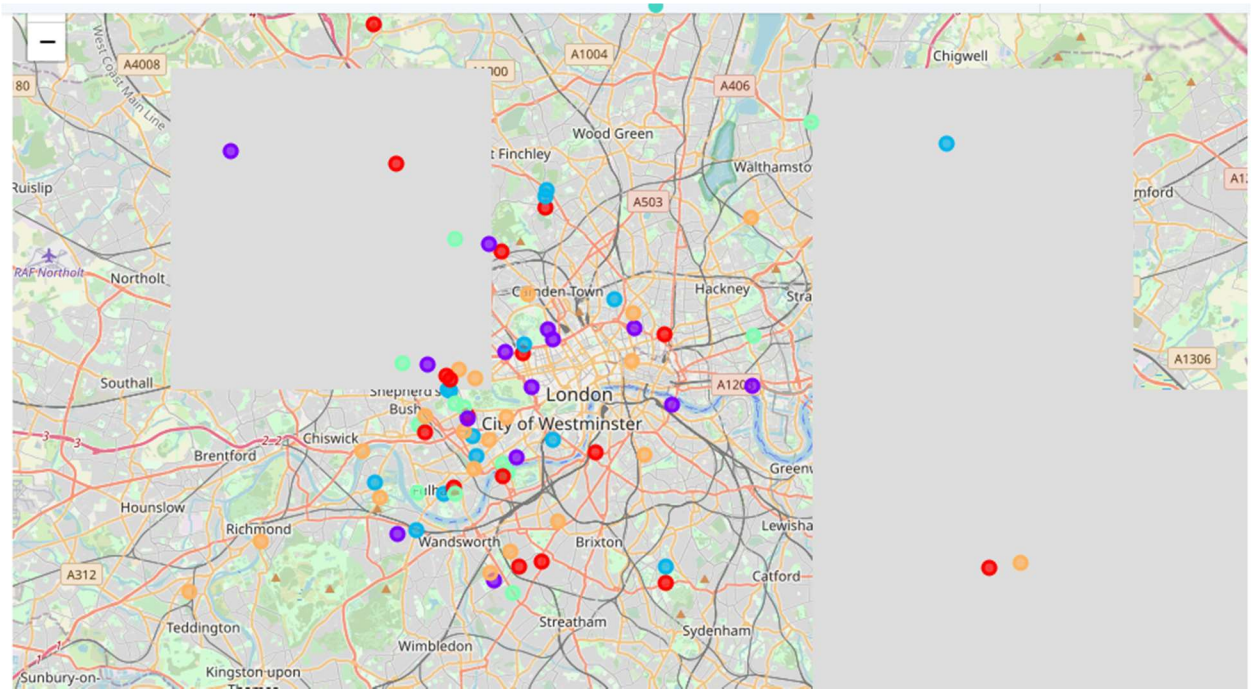
We can now proceed to the Modeling phase. We will analyze neighborhoods to recommend real estates where home buyers can make a real estate investment. We will then recommend profitable venues according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools,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

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.

	Street	Avg_Price	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
196	ALBION SQUARE	2.450000e+06	-41.273758	173.289393	1	Café	Coffee Shop	Bar	Indian Restaurant	Pub	Restaurant
390	ANHALT ROAD	2.435000e+06	51.480316	-0.166801	3	Pub	Grocery Store	Garden	English Restaurant	Plaza	French Restaurant
405	ANSDELL TERRACE	2.250000e+06	51.499890	-0.189103	0	Juice Bar	Hotel	Italian Restaurant	Restaurant	Clothing Store	Indian Restaurant
422	APPLEGARTH ROAD	2.400000e+06	53.748654	-0.326670	3	Pub	Bar	Nightclub	Casino	Factory	Egyptian Restaurant
855	BARONSMEAD ROAD	2.375000e+06	51.477315	-0.239457	2	Food & Drink Shop	Movie Theater	Breakfast Spot	Pub	Park	Indie Movie Theater
981	BEAUCLERC ROAD	2.480000e+06	30.211452	-81.617981	1	Harbor / Marina	Speakeasy	Automotive Shop	Pizza Place	Zoo	Factory
1102	BELVEDERE DRIVE	2.340000e+06	38.072818	-78.458796	2	Pool	Playground	Athletics & Sports	Zoo	Falafel Restaurant	Egyptian Restaurant

Cluster Result



Results and Discussion section

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 under-priced 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 homebuyers clientele in London to make wise and effective decisions. As a result, the business problem we were posing was: how could we provide support to homebuyers clientele 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.