Battle of Neighbourhoods

Valentia

April 2019

1 Introduction

The below business problem will assist us in knowing if we can use real estate historical data to predict which neighbourhoohds are best for starting real estate business. It is known that essential facilities such as elementary schools, high schools, hospitals and grocery stores influence the house prices for a given area. Can we use foursquare API and property data to help home buyers make the right decision.

1.0.1 Business Problem

In this scenario, it is urgent to adopt machine learning tools in order to assist homebuyers 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 homebuyers 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 and grocery stores.

2 Data Section

The following data on London properties and the relative price paid data were extracted from the HM Land Registry (http://landregistry.data.gov.uk/). The 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.

3 Methodology

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

- Collect Data
- Data understanding and Exploration
- Data preparation and pre-processing
- Modelling

4 Results

	Street	Avg_Price	Latitude	Longitude
194	ALBION SQUARE	2.450000e+06	-41.273758	173.289393
388	ANHALT ROAD	2.435000e+06	51.480326	-0.166761
402	ANSDELL TERRACE	2.250000e+06	51.499890	-0.189103
417	APPLEGARTH ROAD	2.400000e+06	53.749244	-0.326780
845	BARONSMEAD ROAD	2.375000e+06	51.477315	-0.239457
969	BEAUCLERC ROAD	2.480000e+06	51.499577	-0.229033
1091	BELVEDERE DRIVE	2.340000e+06	38.201316	-84.623076
1204	BICKENHALL STREET	2.208500e+06	51.521197	-0.158934
1242	BIRCHLANDS AVENUE	2.217000e+06	51.448394	-0.160468
1539	BRAMPTON GROVE	2.456875e+06	51.570365	-0.283394
1617	BRIARDALE GARDENS	2.397132e+06	51.560175	-0.195431
1781	BROOKWAY	2.400000e+06	45.432185	-122.802812
1896	BURBAGE ROAD	2.445000e+06	51.450967	-0.095288
1960	BURY WALK	2.492500e+06	52.145529	-0.423593
2041	CALEDONIAN ROAD	2.395704e+06	51.548482	-0.118359
2046	CALLCOTT STREET	2.375000e+06	51.508350	-0.198328
2107	CAMPDEN HILL ROAD	2.379653e+06	51.506461	-0.198896
2114	CAMPION ROAD	2.461000e+06	52.681351	0.965439
2136	CANNING PLACE	2.425000e+06	51.499570	-0.184248
2203	CARLISLE ROAD	2.200000e+06	42.549099	-71.416669

Figure 1: This table shows the neighbourhoods average by house price.

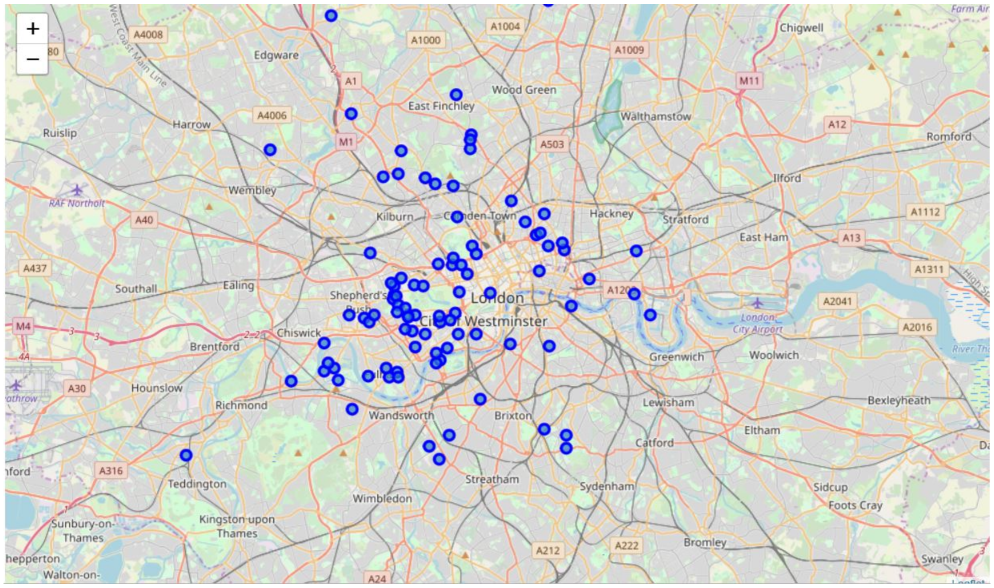

Figure 2: This map shows Map of London with neighbourhoods of interest.

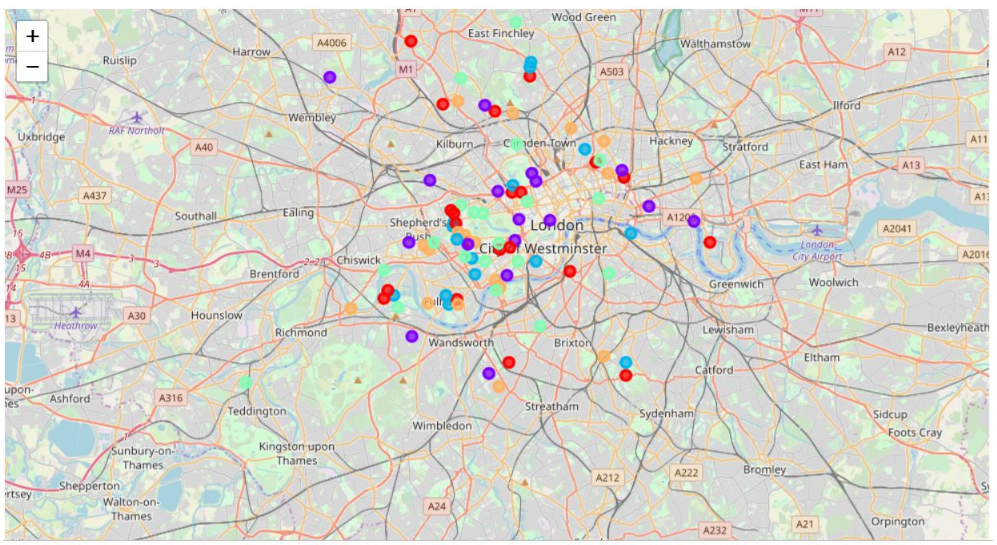

Figure 3: This map shows map of London after K-means clustering technique has been done. For this cluster, K is 5.

5 Results and Discussion

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 and 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.

6 Conclusion

In conclusion, we adopt machine learning tools in order to assist homebuyers 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 in to purchase a suitable real estate in London in this uncertain economic and financial scenario? This algorithm is suppose to help home buyers to make better decisions when purchasing houses.

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.

The Methodology section comprised four stages:

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

In the modeling section, we used the k-means clustering, an unsupervised learning 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.

Even though the London housing market is in the gutter, it is still an "A-okay" for business affairs. Results were discussed under two main perspectives:

1. 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. 2. 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.