The Battle of the Neighborhoods Of London

Adam Bellos

Introduction

Background-Business Problem

London is one of the most famous and impactful cities in the world. It is the capital and the largest city of England and the United Kingdom. London has been called the world's most powerful, most influential, most desirable city to live and specially to work in. London exerts a considerable impact upon arts, commerce, education, health care, novelties, research and development but its most important facet is London's economical structure as it is considered to be a leader and regulator of global economy.

Beyond this timeless attractions of London UK faces one of its most crucial period. As most of you could imagine i am referring to the Brexit consequences. The UK left the EU on 31 January of 2020. UK has already suffered from Brexit .The economy has slowed, and many businesses have moved their headquarters to the EU. Here are some of the impacts on growth, trade, and jobs. There would also be consequences specific to Ireland, London, and Scotland. In this context the purpose of this project is to explore how real estate in London has been affected. The result of this project is aiming to be beneficial for the potential buyers as well for the real estate professionals

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.

Methodology

In this section I am going to delineate the process from the data acquisition to the final result of the project.

The first vital step is to obtain adequate information about the location and the prices of the apartments in London. This information has been derived from landregistry.data.gov.uk Hm Land Registry publishes the following public datasets on GOV.UK:

- Price paid data updated monthly, data available from 1995
- Transaction data updated monthly, data available from 2011
- UK House Price Index downloads updated monthly, data available from 1995.HM Land Registry publish the UK House Price Index on behalf of Office

for National Statistics, Registers of Scotland and Land and Property Services Northern Ireland

Next, I converted the above data into a pandas dataframe

From the above data I took the streets of London and the available prices for apartments on those streets

2018-09-25 00:00	SK7 1AR	D	N	F	5	Unnamed: 8	OAK MEADOW	BRAMHALL	STOCKPORT	STOCKPORT.1	GREATER MANCHESTER	A	A.1
2018-09-24 00:00	M6 8GQ	D	N	F	1	NaN	RIVINGTON ROAD	NaN	SALFORD	SALFORD	GREATER MANCHESTER	A	A
2018-09-28 00:00	WA3 2UE	D	Y	F	35	NaN	STONEACRE CLOSE	LOWTON	WARRINGTON	WIGAN	GREATER MANCHESTER	A	A
2018-08-29 00:00	OL6 6RJ	S	N	F	102	NaN	THORNFIELD GROVE	NaN	ASHTON-UNDER-LYNE	TAMESIDE	GREATER MANCHESTER	A	A
2018-06-15 00:00	M46 0TW	S	Y	F	37	NaN	THREADNEEDLE PLACE	ATHERTON	MANCHESTER	WIGAN	GREATER MANCHESTER	A	A
2018-09-28 00:00	M28 3XS	D	Y	L	9	NaN	MARPLE GARDENS	WORSLEY	MANCHESTER	SALFORD	GREATER MANCHESTER	A	A

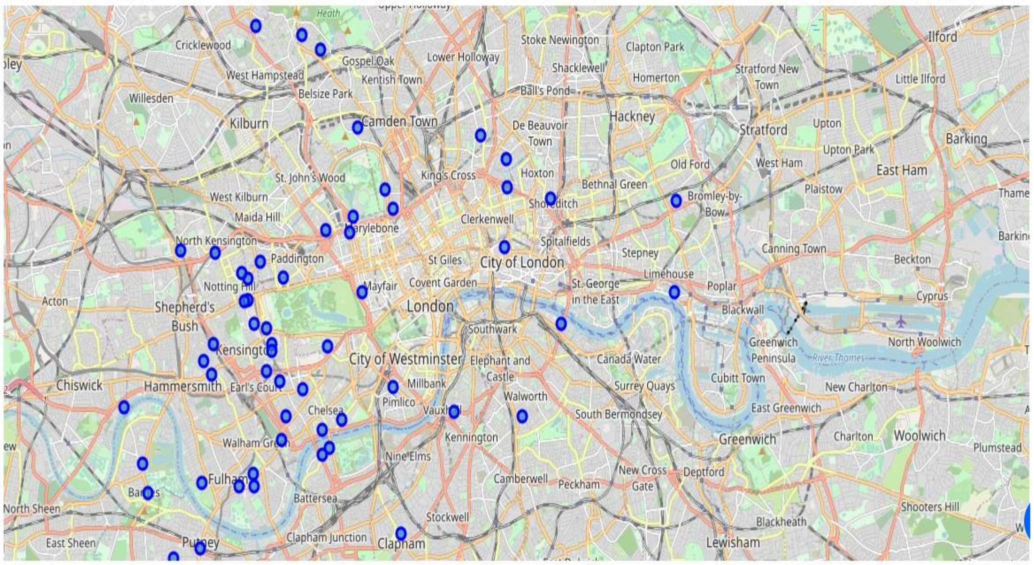
Subsequently I grouped the apartments based on the street they are located and calculated the average price of an apartment per street . The next step is to input the budget of the potential buyer in order to sort the avaliable choices. Here is a segment of this information

	Street	Avg_Price
196	ALBION SQUARE	2.450000e+06
390	ANHALT ROAD	2.435000e+06
405	ANSDELL TERRACE	2.250000e+06
422	APPLEGARTH ROAD	2.400000e+06
855	BARONSMEAD ROAD	2.375000e+06
981	BEAUCLERC ROAD	2.480000e+06
1102	BELVEDERE DRIVE	2.340000e+06
1215	BICKENHALL STREET	2.208500e+06
1253	BIRCHLANDS AVENUE	2.217000e+06
1553	BRAMPTON GROVE	2.456875e+06
1632	BRIARDALE GARDENS	2.397132e+06
1797	BROOKWAY	2.400000e+06
1914	BURBAGE ROAD	2.445000e+06
1980	BURY WALK	2.492500e+06
2068	CALLCOTT STREET	2.375000e+06
2129	CAMPDEN HILL ROAD	2.379653e+06
2136	CAMPION ROAD	2.461000e+06
2158	CANNING PLACE	2.425000e+06
2225	CARLISLE ROAD	2.200000e+06
2230	CARLTON GARDENS	2.483500e+06
[illegible]	CARLYLE COURT	[illegible]

Then using geolocator we find the coordinates of each street and we expand the above dataframe with a column consisting of those coordinates.

	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
2129	CAMPDEN HILL ROAD	2.379653e+06	51.508111	-0.199667
2136	CAMPION ROAD	2.461000e+06	52.681078	0.965599

Taking leverage, again, of geolocator we get the coordinates of the center of London in order to create a map using folium library



Use of Foursquare

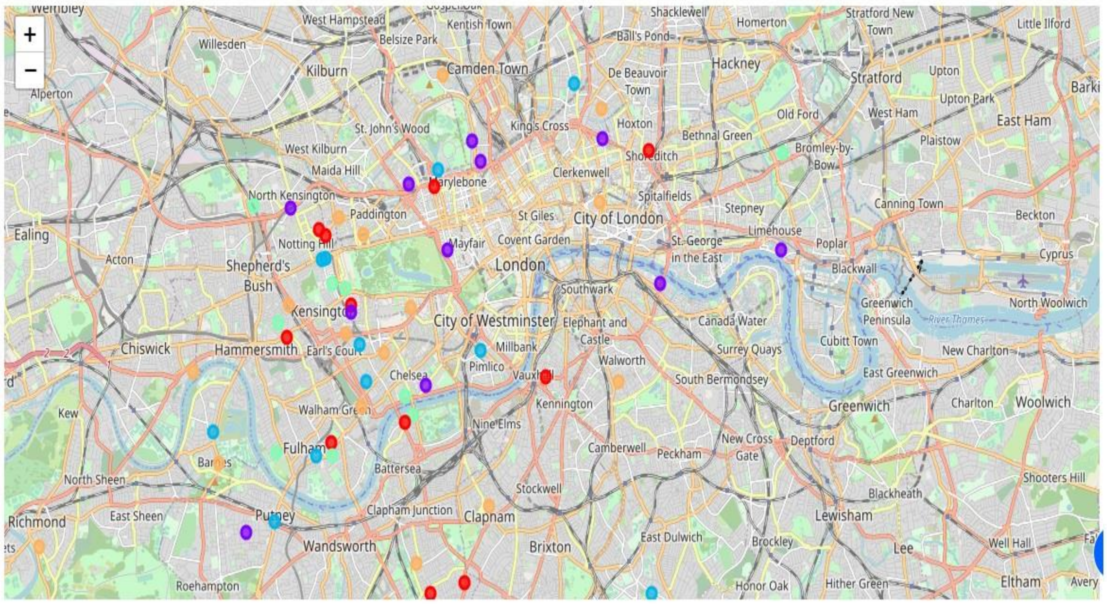
I created a function which given a name of an area and its coordinates is returning the around venues in a radius of 500 meters in a dataframe format. The coding of this function uses Foursquare api query techniques. This function has been created in order to be called by giving it as parameters the components of our dataset. Specifically the names London's streets and their coordinates which we have obtained above.

	Street	Street Latitude	Street Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ALBION SQUARE	-41.273758	173.289393	The Free House	-41.273340	173.287364	Bar
1	ALBION SQUARE	-41.273758	173.289393	The Indian Cafe	-41.273308	173.286530	Indian Restaurant
2	ALBION SQUARE	-41.273758	173.289393	Queen's Gardens	-41.273671	173.291383	Park
3	ALBION SQUARE	-41.273758	173.289393	Urban	-41.274355	173.286317	New American Restaurant
4	ALBION SQUARE	-41.273758	173.289393	Fish Stop	-41.276010	173.289592	Fish & Chips Shop
5	ALBION SQUARE	-41.273758	173.289393	Deville Cafe	-41.271941	173.285535	Beer Garden
6	ALBION SQUARE	-41.273758	173.289393	Fresh Choice	-41.272194	173.287218	Supermarket
7	ALBION SQUARE	-41.273758	173.289393	The Bridge Street Collective	-41.272520	173.285517	Café
8	ALBION SQUARE	-41.273758	173.289393	Mango	-41.274460	173.285345	Indian Restaurant
9	ALBION SQUARE	-41.273758	173.289393	Hopgood's	-41.274749	173.283831	Restaurant
10	ALBION SQUARE	-41.273758	173.289393	The Vic Mac's Brew Bar	-41.274757	173.283914	Pub
11	ALBION SQUARE	-41.273758	173.289393	The Kitchen	-41.272360	173.285500	Café
12	ALBION SQUARE	-41.273758	173.289393	cod and lobster	-41.275203	173.283747	Seafood Restaurant
13	ALBION SQUARE	-41.273758	173.289393	Burger Culture	-41.274750	173.284030	Burger Joint
14	ALBION SQUARE	-41.273758	173.289393	Sprig & Fern	-41.274508	173.286527	Brewery
15	ALBION SQUARE	-41.273758	173.289393	Columbus Coffee	-41.274759	173.285391	Coffee Shop
16	ALBION SQUARE	-41.273758	173.289393	La Gourmandise	-41.274262	173.286211	French Restaurant
17	ALBION SQUARE	-41.273758	173.289393	Lambretta's Cafe & Bar	-41.274272	173.284462	Café

Now that we have our venues it's time to explore them using python methods. So, based on our search 4255 venues have been found which can be classified in 355 unique categories. Also we searched the found the top 5 venues/facilities per street that may be profitable real estate investments.In order to get a better image I created a dataframe   which shows the top 10 venue categories per street.

	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	Indian Restaurant	Restaurant	Pub	Bar	Coffee Shop	Burger Joint	French Restaurant	New American Restaurant	Fish & Chips Shop
1	ANHALT ROAD	Pub	Grocery Store	Plaza	French Restaurant	Gym / Fitness Center	Japanese Restaurant	Cocktail Bar	English Restaurant	Garden	Diner
2	ANSDELL TERRACE	Hotel	Italian Restaurant	Juice Bar	Clothing Store	Restaurant	Bakery	Garden	Pub	Indian Restaurant	French Restaurant
3	APPLEGARTH ROAD	Pub	Nightclub	Bar	Casino	Flower Shop	Farmers Market	Event Space	Food & Drink Shop	Exhibit	Factory
4	BARONSMEAD ROAD	Food & Drink Shop	Coffee Shop	Breakfast Spot	Sports Club	Nature Preserve	Café	Farmers Market	Thai Restaurant	Park	Restaurant

Finally I used an unsupervised machine learning process called Kmeans in order to cluster my dataset and merged the outcome with the data by expanading the dataframe



Results

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.