The Battle of Neighborhoods
London Housing Market 2019 (Jan-May)

By Xiaoyan (Elva) Ma

Content

Introduction

1. Background description

Home price in London growth has been gradually slowing ever since the summer of 2016. A moribund market is preventing potential sellers from bringing their homes to the market and an ongoing decline in new instructions being listed for sale has intensified of late. A recent survey of U.K. residential property from the Royal Institution of Chartered Surveyors (RICS) concluded that Brexit is currently the main obstacle for market activity. And uncertainty over how the U.K. leaves the European Union was holding back both buyers and sellers of property. House prices may be continuing to fall with the number of new inquiries down for the eighth consecutive month,

Introduction

2. Business Problem

Suppose I am a professional real estate trading consultant, my clients come to my office and ask me how can I assist them to purchase a suitable real estate in London in this uncertain economic? What they can put their money on?

Data Collection

Price Paid Data includes information on all property sales in England and Wales that are sold for full market value and are lodged with registration. The files include standard and additional price paid data transactions received at HM Land Registry in the period from January 1st to May 31st in 2019 at https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads.

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.

Methodology

To explore and target recommended locations across different venues according to the presence of amenities and essential facilities, I 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, I will be able to recommend profitable real estate investments.

The necessary libraries will be used in this processing are: os, numpy, pandas, datetime, json, Nominatim, requests, matplotlib.cm, and folium.

Methodology

1. Data preprocessing

This dataset consists is over 211899 rows and 16 columns. The original data without the headers cannot be process, so first I created a header for each column as 'TUID', 'Sold_Price', 'Date', 'Postcode', 'Prop_Type', 'Old_New', 'Duration', 'PAON', 'SAON', 'Street', 'Locality_Name', 'Town_City', 'District', 'County', 'PPD_Cat_Type' and 'Record_Status'.

	TUID	Sold_Price	Date	Postcode	Prop_Type	Old_New	Duration	PAON	SAON	Street	Locality_Name	Town_City
0	{87E1551E-C274-6405-E053-6C04A8C0B2EE}	119000	2019-04-12 00:00	SN2 2HZ	F	N	L	53	NaN	ROSE STREET	NaN	SWINDON
1	{87E1551E-C275-6405-E053-6C04A8C0B2EE}	209000	2019-04-05 00:00	SN5 5GP	T	N	F	31	NaN	DANESTONE CLOSE	MIDDLELEAZE	SWINDON
2	{87E1551E-C276-6405-E053-6C04A8C0B2EE}	213000	2019-03-29 00:00	SN5 4AZ	T	N	F	72	NaN	REIDS PIECE	PURTON	SWINDON
3	{87E1551E-C277-6405-E053-6C04A8C0B2EE}	250000	2019-04-03 00:00	SN5 5FP	D	N	F	5	NaN	NEVIS CLOSE	SPARCELLS	SWINDON

Methodology

2. Explore and Understand Data

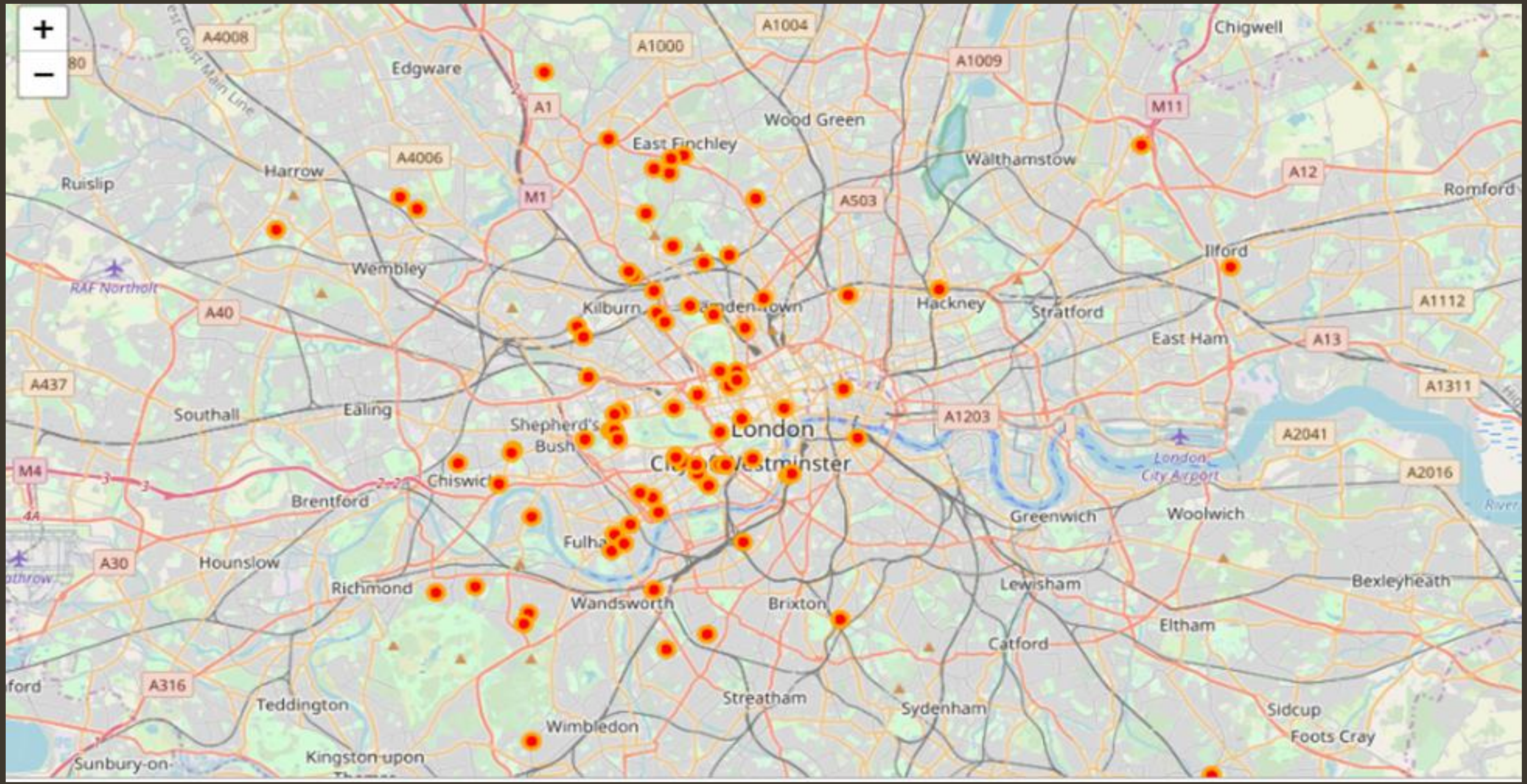
I need to format the date column and sort by date of sale, make a list of street names in London city and give meaningful names to the columns as Street and Average Price. In this case I limited the average price purchase power from 1 million to 2.8 million.

Pair of IP address to geographical location to provide geolocation data of Latitude and Longitude for each street

Create map of London using latitude and longitude values and add marker to map. The print show on below map in color red and orange are the target properties that match our request as the average price is between 1 million to 2.8 million. And I can see most they are in West London, South West and North West London

Methodology

2. Explore and Understand Data



Methodology

3. Modeling

In this step I will explore the first neighborhood into the data frame, first I create a function to process to all the streets I target in London and establish a new dataframe with the 124 rows and 6 columns

I analyze each street by add street column back to dataframe and fixed it. I got 123 rows and 331 columns of streets grouped in London. And found out each street along with the top 6 most common venues

	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
0	AIREDALE AVENUE	Pub	Italian Restaurant	Pizza Place	Hotel	Burger Joint	Thai Restaurant
1	ALBERT EMBANKMENT	Park	Café	Pub	Hotel	Coffee Shop	Restaurant
2	ALBION DRIVE	Rock Club	Supermarket	Indian Restaurant	Fast Food Restaurant	Zoo Exhibit	Food Stand
3	ALBION SQUARE	Café	Pub	Bar	Indian Restaurant	Restaurant	Coffee Shop
4	ALDERSGATE STREET	Italian Restaurant	Coffee Shop	Hotel	French Restaurant	Gym / Fitness Center	Plaza

Methodology

4. Clustering

In this step I start to cluster the properties by venues, facukutues, amenitites nearby the most profitable real estate investments in London. First I set number of clusters is 5, run K-means clustering and check cluster labels generated for each row in the dataframe. For example,

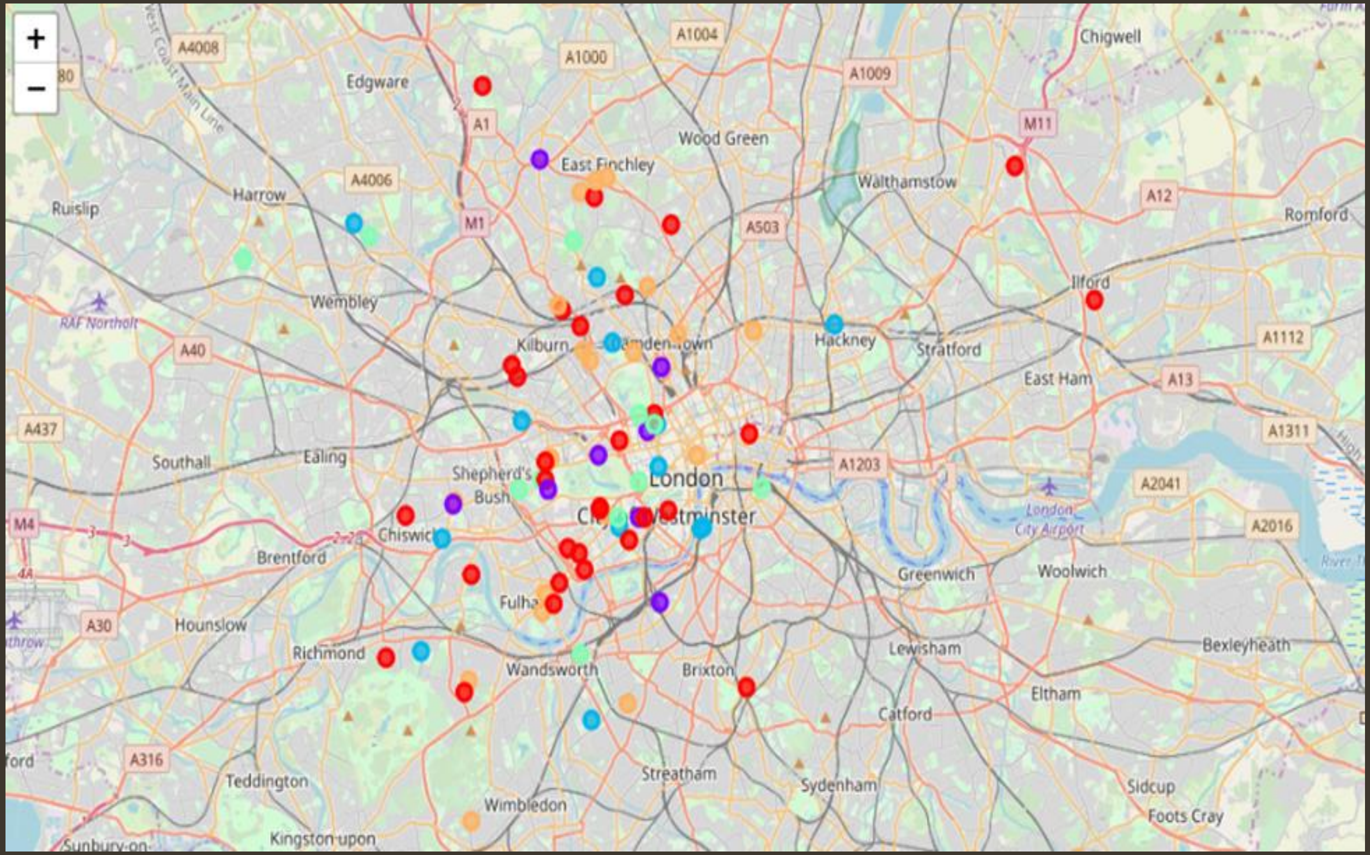
Aberdeen Park is label in cluster 2, Albion Drive is label in cluster 0 and Albion Square is label in cluster 4

	Street	Avg_Price	Latitude	Longitude	Cluster Labels
19	ABERDEEN PARK	2100000.0	35.928397	-86.519993	2
74	AIREDALE AVENUE	2025000.0	51.491383	-0.248097	2
87	ALBERT EMBANKMENT	2020000.0	51.493949	-0.121144	2
92	ALBION DRIVE	2180000.0	52.653656	1.312262	0
96	ALBION SQUARE	2450000.0	-41.273758	173.289393	4

Methodology

4. Clustering

And all cluster can be display on below map that 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



Methodology

4. Clustering

And each cluster can also show in dataframe as following:

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
92	2180000.0	Rock Club	Supermarket	Indian Restaurant	Fast Food Restaurant	Zoo Exhibit	Food Stand
105	2307473.0	Italian Restaurant	Coffee Shop	Hotel	French Restaurant	Gym / Fitness Center	Plaza
858	2200000.0	Hotel	Middle Eastern Restaurant	Italian Restaurant	Sandwich Place	Lebanese Restaurant	Coffee Shop
868	2240000.0	Hotel	Coffee Shop	Sandwich Place	Theater	Sushi Restaurant	Gym / Fitness Center
940	2310000.0	Liquor Store	Gastropub	Seafood Restaurant	Indian Restaurant	Food Stand	Fish & Chips Shop

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
255	2.600000e+06	Grocery Store	Pub	Coffee Shop	Park	Indian Restaurant	Japanese Restaurant
428	2.600000e+06	NaN	NaN	NaN	NaN	NaN	NaN
976	2.568800e+06	Pub	Coffee Shop	Pizza Place	Greek Restaurant	Supermarket	Music Venue
1919	2.535000e+06	Italian Restaurant	Restaurant	Café	Hotel	Pub	Coffee Shop
2345	2.623333e+06	Paper / Office Supplies Store	Zoo Exhibit	Film Studio	Fish Market	Flea Market	Food

Methodology

4. Clustering

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
19	2100000.0	NaN	NaN	NaN	NaN	NaN	NaN
74	2025000.0	Pub	Italian Restaurant	Pizza Place	Hotel	Burger Joint	Thai Restaurant
87	2020000.0	Park	Café	Pub	Hotel	Coffee Shop	Restaurant
643	2094150.0	Coffee Shop	Restaurant	Café	Italian Restaurant	Pizza Place	Mediterranean Restaurant
942	2000000.0	Café	Italian Restaurant	Hotel	Boutique	Japanese Restaurant	Restaurant

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
652	2800000.0	Coffee Shop	Pub	Restaurant	Italian Restaurant	Food Truck	Hotel
713	2800000.0	Men's Store	Lake	Middle Eastern Restaurant	Zoo Exhibit	Food Stand	Fish & Chips Shop
988	2675000.0	Pub	Coffee Shop	Bakery	Yoga Studio	Indian Restaurant	Hotel
1218	2750000.0	Gastropub	Construction & Landscaping	Zoo Exhibit	Food Truck	Fish Market	Flea Market
1310	2800000.0	Bar	Lounge	Pizza Place	Gay Bar	Mexican Restaurant	Theater

Methodology

4. Clustering

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
96	2450000.0	Café	Pub	Bar	Indian Restaurant	Restaurant	Coffee Shop
254	2425000.0	Pub	Grocery Store	Coffee Shop	Indian Restaurant	Japanese Restaurant	Auto Garage
454	2450000.0	NaN	NaN	NaN	NaN	NaN	NaN
466	2500000.0	Restaurant	Performing Arts Venue	Fast Food Restaurant	Fish & Chips Shop	Fish Market	Flea Market
808	2475000.0	Café	Italian Restaurant	Coffee Shop	Hotel	Exhibit	Boutique

Conclusion

After processing all the above data and I can summaries and recommend that 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. I also clustered 5 area that

Cluster 0: 1st most common venue in are club, restaurant hotel and liquor store, this cluster will be highly recommended for our client who want to invest for business traveler or sightseeing;

Cluster 1, 2 and 4: most common venue in those clusters are grocery stores, Pub, restaurants, cafe and supermarkets. Those area target home buyers prone to invest the residential live, people living in this area will be very convenient for shopping everything they need.

Cluster 3: most common venue are coffee shop, men’s store, pub, and bar. This cluster will be called recreation area, it can be targeted for people who enjoy relaxation life after work.