Capstone project-Battle of quarters

Business problems section

Background

According to Bloomberg News, the London housing market is in a rut. It is currently facing a number of different headwinds, including the prospect of higher taxes and a warning from the Bank of England that the cost of housing in the UK could fall by as much as 30 percent in the event of a disorderly exit from the European Union. More specifically, the four overlooked cracks suggest that the London market may be in worse shape than many assume: hidden price falls, record low sales, an Exodus of Housebuilders and tax hikes targeting overseas home buyers in England and Wales.

business problem

In this case, there is an urgent need to adopt machine learning tools to help London home buyers make wise and effective decisions. As a result, the business problem we are currently facing is the following: how could we support homebuyers clients in purchasing suitable properties in London in this uncertain economic and financial scenario? To solve this business problem, we are going to bring together London boroughs to recommend locations and the current average property price where homebuyers can make a property investment. We will recommend favorable locations according to the amenities and basic amenities surrounding places such as elementary schools, high schools, hospitals, and grocery stores.

Data section

Data on London properties and the relative price paid were extracted from the HM land registry (http://landregistry.data.gov.uk/). the following fields contain address data included in the paid price data: postal code; name of the main address object paon. This is usually a house number or name; caon is a secondary address object. If there is an extension, for example, the building is divided into apartments, then there will be a district; street; locality; city/city; district; district.

To explore and navigate recommended locations in various locations depending on the availability of amenities and basic amenities, we will access data through the Foursquare API and organize it as a data frame for visualization. By combining data on London properties and the relative price paid from the HM land registry, as well as data on amenities and major properties surrounding such properties from the Foursquare API, we can recommend profitable real estate investments.

2. Explore and Understand Data

```
[6]: df_ppd.head(5)
```

[6]:

	{79A74E21-D11E-1289-E053-6B04A8C01627}	770000	2018-09-25 00:00	SK7 1AR	D	N	F	5	Unnamed: 8	OAK MEADOW	BRAMHALL	STOCKPORT	STOCKPORT.1	GREATER MANCHESTER	A	A.1
0	{79A74E21-D11F-1289-E053-6B04A8C01627}	253500	2018-09-24 00:00	M6 8GQ	D	N	F	1	NaN	RIVINGTON ROAD	NaN	SALFORD	SALFORD	GREATER MANCHESTER	A	A
1	{79A74E21-D120-1289-E053-6B04A8C01627}	231950	2018-09-28 00:00	WA3 2UE	D	Y	F	35	NaN	STONEACRE CLOSE	LOWTON	WARRINGTON	WIGAN	GREATER MANCHESTER	A	A
2	{79A74E21-D121-1289-E053-6B04A8C01627}	112500	2018-08-29 00:00	OL6 6RJ	S	N	F	102	NaN	THORNFIELD GROVE	NaN	ASHTON-UNDER-LYNE	TAMESIDE	GREATER MANCHESTER	A	A
3	{79A74E21-D122-1289-E053-6B04A8C01627}	184995	2018-06-15 00:00	M46 0TW	S	Y	F	37	NaN	THREADNEEDLE PLACE	ATHERTON	MANCHESTER	WIGAN	GREATER MANCHESTER	A	A
4	{79A74E21-D123-1289-E053-6B04A8C01627}	214995	2018-09-28 00:00	M28 3XS	D	Y	L	9	NaN	MARPLE GARDENS	WORSLEY	MANCHESTER	SALFORD	GREATER MANCHESTER	A	A

```
[7]: df_ppd.shape
```

```
[7]: (1029228, 16)
```

3. Data preparation and preprocessing

At this stage, we prepare our dataset for the modeling process, opting for the most suitable machine learning algorithm for our scope. Accordingly, we perform the following steps:

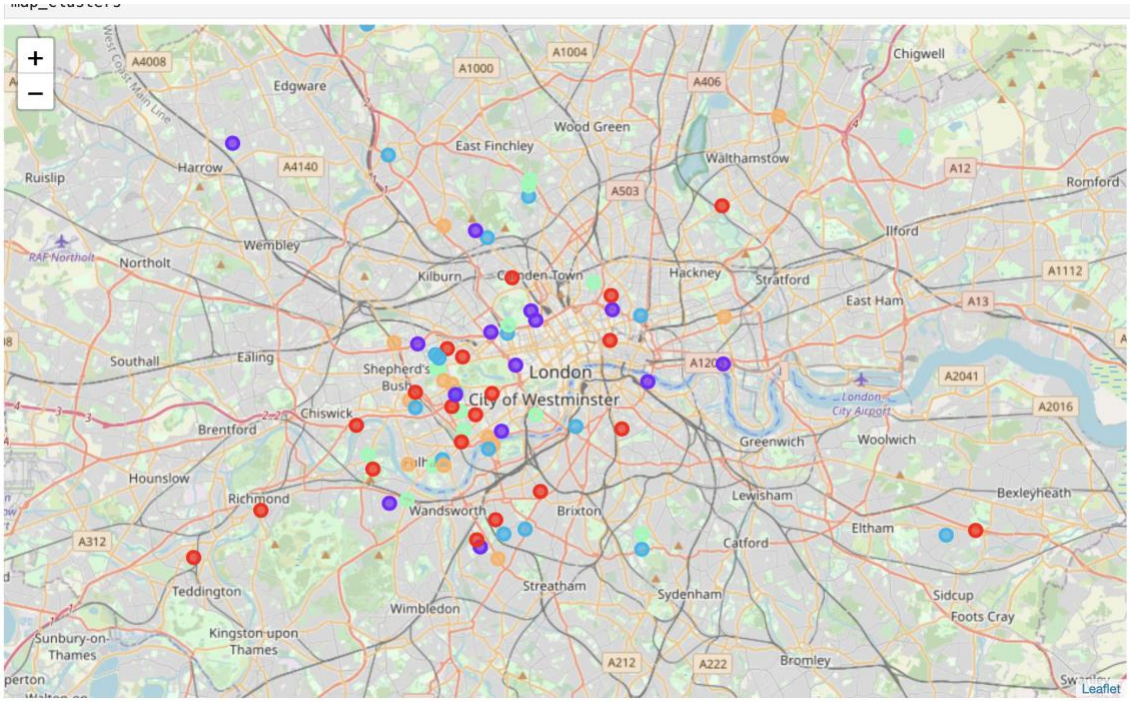
- Rename the column names
- Format the date column
- Sort data by date of sale
- Select data only for the city of London
- Make a list of street names in London
- Calculate the street-wise average price of the property
- Read the street-wise coordinates into a data frame, eliminating recurring word London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

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.

K-Means clustering

49]:



Results and discussion section

First of all, despite the fact that the London housing market may be in a rut, it still remains "forever green" for business.

We can discuss our results from two main points of view.

First, we can research them according to London boroughs / districts. It is interesting to note that while West London (Notting hill, Kensington, Chelsea, Marylebone) and North West London (Hampstead) can be considered very profitable places to buy property according to the amenities and basic amenities surrounding places such as primary schools, secondary schools, hospitals and grocery stores, South West London (Wandsworth, Balham) and North West London (Islington) are emerging as the next future elite locations with a wide range of amenities and amenities. Accordingly, it would be possible to focus on low property prices in these areas of London to make a business case.

Second, we can analyze our results according to the five clusters we have created. Despite the fact that all clusters could praise the optimal set of facilities and amenities, we found two main patterns. The first model we are talking about, i.e. clusters 0, 2 and 4, can be aimed at home buyers who tend to live in "green" areas with parks and embankments. Instead, the second model we are talking about, i.e. clusters 1 and 3, can be aimed at people who like pubs, theaters and football.

Conclusion

To sum up, according to Bloomberg News, the London housing market is in a rut. It is currently facing a number of different headwinds, including the prospect of higher taxes and a warning from the Bank of England that the cost of housing in the UK could fall by as much as 30 percent in the event of a disorderly exit from the European Union. In this case, there is an urgent need to adopt machine learning tools to help London home buyers make wise and effective decisions. As a result, the business challenge we set for ourselves was: how could we support homebuyers clients in purchasing suitable properties in London in this uncertain economic and financial scenario?

To solve this business problem, we have grouped London boroughs to recommend locations and the current average property price where homebuyers can make property investments. We recommended favorable locations according to the amenities and basic amenities surrounding places such as elementary schools, high schools, hospitals, and grocery stores.

First, we collected data on London real estate, and the relative price data was extracted from the HM land registry (http://landregistry.data.gov.ukin addition, to explore and navigate the recommended locations in various locations depending on the availability of amenities and major facilities, we accessed the data via the Foursquare API and arranged it as a data frame for visualization. By combining data on London properties and the relative price paid from the HM land registry, as well as data on amenities and major properties surrounding such properties from the Foursquare API, we were able to recommend profitable real estate investments.

Second, the methodology section consisted of four stages: 1. collecting inspection data; 2. Researching and understanding data; 3.Preparing and pre-processing data; 4.Modeling. In particular, in the modeling section, we used the K-means clustering method, because it is fast and efficient in terms of computational costs, very flexible for accounting for mutations in the London property market, and accurate.

Finally, we concluded that while the London housing market may be in a rut, it is still "forever green" for business. We discussed our results from two main points of view. First, we looked at them according to London boroughs / districts. although West London (Notting hill, Kensington, Chelsea, Marylebone) and North-West London (Hampstead) can be considered very profitable places to buy property according to the amenities and basic amenities surrounding such places, i.e. primary schools, secondary schools, hospitals and grocery stores, South West London (Wandsworth, Balham) and North West London (Islington) are becoming the next future luxury properties with a wide range of amenities and amenities. Accordingly, it would be possible to focus on low property prices in these areas of London to make a business case. Second, we analyzed our results for the five clusters we created. While clusters 0, 2, and 4 may target home buyers who tend to live in "green" areas with parks, embankments, clusters 1 and 3 may target people who like pubs, theaters, and soccer.