Explore different Boroughs in London

Lin Yang

10/01/2019

So many Boroughs in London, what are they like?
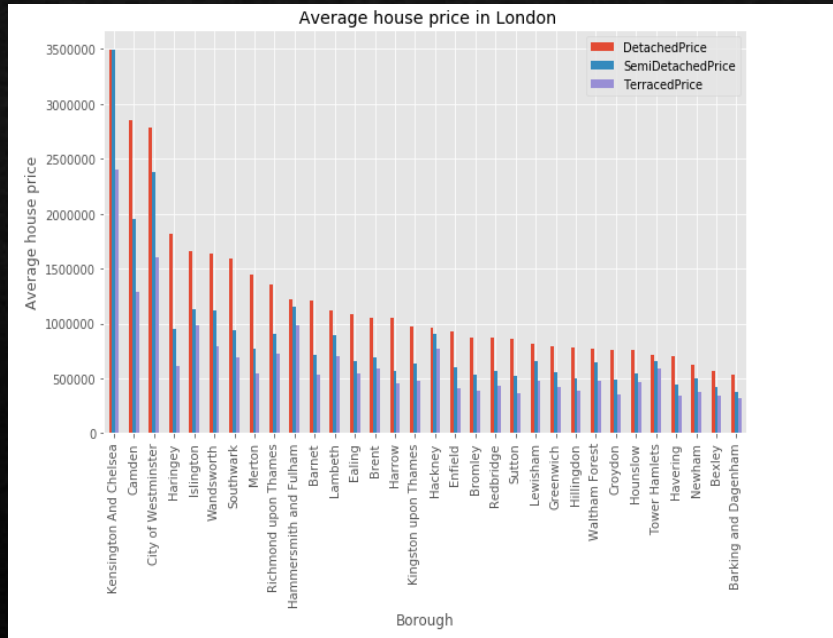
◈ Imagine you are the first time comer to London, where should you live? What are all the areas like ? Are they expensive? Are they safe?

◈ With all these questions in mind, let's have a look at what are the data available to tell us some insights about London....

  ◈ There are 33 Boroughs in London, and

  ◈ We find house price data for all of them, and

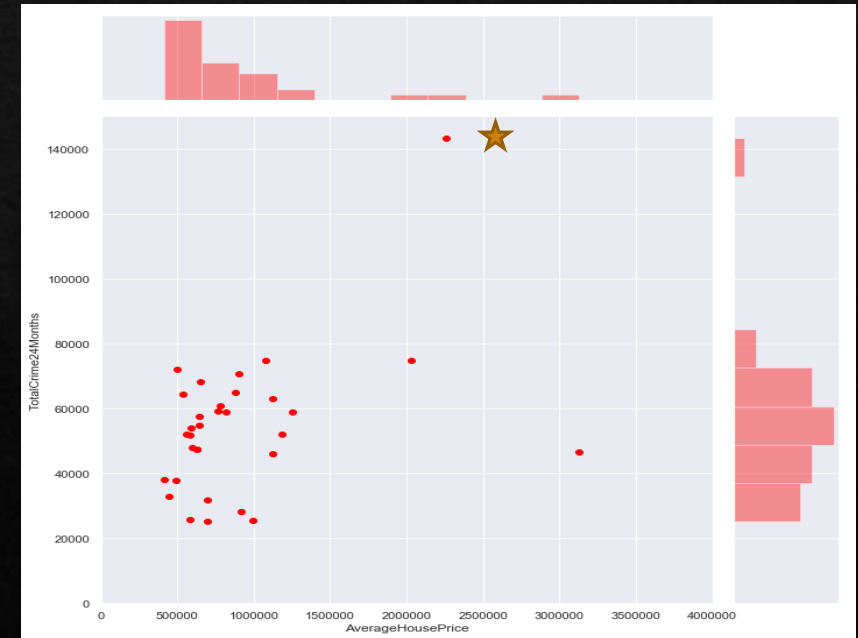 ◈ Both crime data and education data are also ready to use......

Descriptive Statistics

◇ How expensive of houses in London?

◇ Kensington and Chelsea, Camden, City of Westminster can have house price that 5 times more than all other areas, using detached house as example : the 3 area have average price of 2.5 to 3.5 million pounds, where average price of other area of London is 0.5 to 1.0 million pounds.
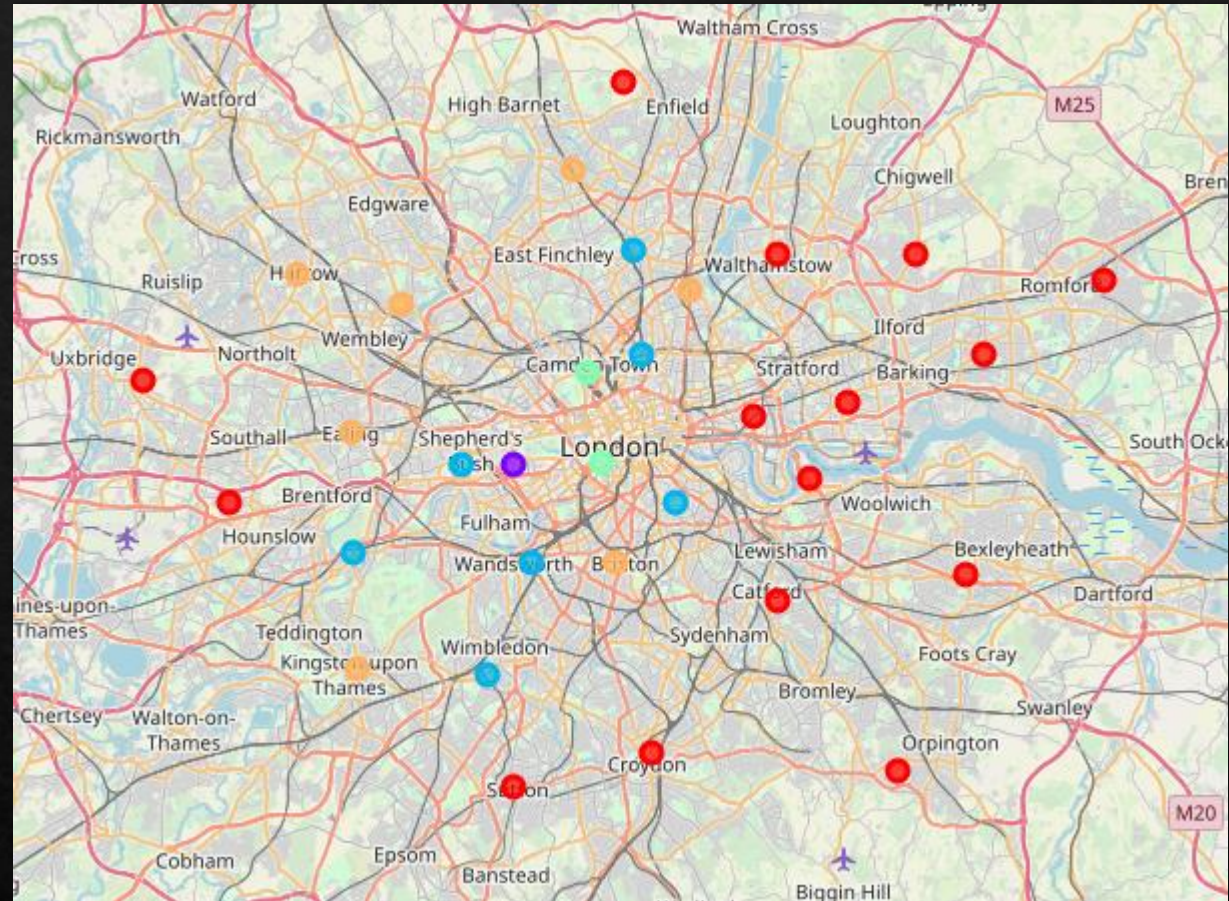
◇ How safe is London?

◇ It is very interesting to see that some high price area (average price of house 2.0 to 2.5 million pounds) are actually have the highest reported crime. A sensible explanation could be rich area would actually encourage more crimes to take risk.

So how can I make my life easier and ask machine to group all the regions to only a few groups ?

- **K-means clustering** is utilised in the analysis, we feed the machine a few set of features about the area (such as price, crime rate, education), and ask machine to provide us a map on the right.

- 5 colours of dots are provided by the machine. In a nutshell, if you chose to live in a red dot in west London, it will be quite similar to any other red dots on the map!

- However, chose to live at a Red dot , it can be quite different from live at a Blue dot.

Conclusion and further recommendation

It doesn't mean Blue dots are definitely better than Red dots! They are just Different!

Certainly you would like more than just price, crime and education to decide where you want to live….

Is my data really tell me the whole story?

They are just different, because machine learning clustering is an unsupervised process, it is up to human to decide how to interpret the result, and tell machine what you would prefer!

Be mindful on the features you can chose to feed the machine, maybe you like more than 5 clusters, maybe we you want to consider how many grocery store or yoga studios are there.

Be curious, if a Brough have high crime rate but most of them are robbery, and another area has low crime rate but all of them are murder, would the crime rate be the ultimate parameter to use?

Appendix

Cluster 0

	Borough	DetachedPrice	SemiDetachedPrice	TerracedPrice	total crime last 24 month	State-funded nursery	State-funded primary	State-funded secondary	Special schools	Independent	All schools	latitude	longitude	Cluster Labels
0	Barking and Dagenham	538728.815971	377073.153252	319039.078555	38231.0	0.0	44.0	13.0	2.0	5.0	65.0	51.546483	0.129350	0
2	Bexley	571584.413457	426410.955665	340698.307180	33099.0	0.0	57.0	16.0	5.0	5.0	84.0	51.451902	0.117179	0
4	Bromley	873546.185963	529320.575549	392514.148248	47973.0	0.0	77.0	19.0	4.0	19.0	122.0	51.367970	0.070062	0
8	Croydon	759048.276654	487158.067247	356658.193863	64392.0	5.0	86.0	25.0	6.0	21.0	144.0	51.376165	-0.098234	0
10	Enfield	930159.555061	596659.126418	408701.544361	57762.0	0.0	68.0	21.0	6.0	13.0	109.0	51.662291	-0.118065	0
11	Greenwich	793966.198834	560556.586051	421058.084906	54167.0	4.0	63.0	17.0	4.0	13.0	102.0	51.493367	0.009821	0
16	Havering	697922.745983	439848.911651	344801.333918	37853.0	0.0	61.0	18.0	3.0	8.0	91.0	51.577924	0.212083	0
17	Hillingdon	785044.786247	499075.210711	388245.840768	52096.0	1.0	69.0	22.0	9.0	9.0	111.0	51.535183	-0.448138	0
18	Hounslow	755847.867062	540223.426334	461322.268024	51863.0	0.0	53.0	20.0	5.0	13.0	92.0	51.482836	-0.388206	0
23	Lewisham	809687.729888	653849.275829	475823.287461	54853.0	2.0	63.0	14.0	5.0	10.0	95.0	51.441458	-0.011701	0
25	Newham	620601.512866	496778.365221	378831.650913	72057.0	7.0	66.0	23.0	2.0	8.0	109.0	51.525516	0.035216	0
26	Redbridge	868358.845533	570406.537589	437624.651014	47511.0	0.0	52.0	19.0	4.0	17.0	95.0	51.588612	0.082398	0
29	Sutton	859262.749785	525741.783393	367987.496668	25766.0	2.0	40.0	15.0	5.0	7.0	71.0	51.361428	-0.193961	0
30	Tower Hamlets	709958.925230	657957.325652	589825.633891	68210.0	6.0	70.0	20.0	6.0	15.0	119.0	51.520261	-0.029340	0
31	Waltham Forest	765047.365607	649242.913465	476564.785896	47579.0	3.0	51.0	18.0	4.0	6.0	85.0	51.588638	-0.011763	0

Cluster 1

	Borough	DetachedPrice	SemiDetachedPrice	TerracedPrice	total crime last 24 month	State-funded nursery	State-funded primary	State-funded secondary	Special schools	Independent	All schools	latitude	longitude	Cluster Labels
20	Kensington And Chelsea	3.494937e+06	3.489462e+06	2.396601e+06	46637.0	4.0	27.0	6.0	2.0	38.0	78.0	51.49908	-0.193825	1

Appendix

Cluster 2

	Borough	DetachedPrice	SemiDetachedPrice	TerracedPrice	total crime last 24 month	State-funded nursery	State-funded primary	State-funded secondary	Special schools	Independent	All schools	latitude	longitude	Cluster Labels
13	Hammersmith and Fulham	1.215054e+06	1.154776e+06	988332.396183	46023.0	4.0	37.0	13.0	5.0	21.0	83.0	51.499016	-0.229150	2
14	Haringey	1.815304e+06	9.448097e+05	615613.119689	63013.0	3.0	64.0	14.0	6.0	15.0	104.0	51.590611	-0.110971	2
19	Islington	1.656366e+06	1.126603e+06	980549.474928	58947.0	3.0	46.0	11.0	6.0	8.0	78.0	51.546506	-0.105806	2
24	Merton	1.441685e+06	7.743666e+05	539612.555855	28389.0	0.0	44.0	9.0	3.0	16.0	73.0	51.409774	-0.210808	2
27	Richmond upon Thames	1.352620e+06	9.068283e+05	726477.845063	25471.0	1.0	45.0	11.0	2.0	23.0	82.0	51.461305	-0.303771	2
28	Southwark	1.586562e+06	9.438556e+05	692751.561823	74980.0	5.0	74.0	19.0	9.0	12.0	120.0	51.483448	-0.082088	2
32	Wandsworth	1.633030e+06	1.122656e+06	793280.378173	52162.0	3.0	61.0	11.0	7.0	31.0	116.0	51.457072	-0.181782	2

Cluster 3

	Borough	DetachedPrice	SemiDetachedPrice	TerracedPrice	total crime last 24 month	State-funded nursery	State-funded primary	State-funded secondary	Special schools	Independent	All schools	latitude	longitude	Cluster Labels
5	Camden	2.850226e+06	1.952578e+06	1.286210e+06	74864.0	1.0	42.0	10.0	5.0	32.0	94.0	51.539026	-0.142552	3
7	City of Westminster	2.783676e+06	2.384244e+06	1.603275e+06	143349.0	4.0	40.0	13.0	3.0	32.0	93.0	51.500175	-0.133233	3

Cluster 4

	Borough	DetachedPrice	SemiDetachedPrice	TerracedPrice	total crime last 24 month	State-funded nursery	State-funded primary	State-funded secondary	Special schools	Independent	All schools	latitude	longitude	Cluster Labels
1	Barnet	1.213280e+06	714684.582475	536867.328748	59112.0	4.0	91.0	26.0	6.0	37.0	166.0	51.625149	-0.152936	4
3	Brent	1.055258e+06	695558.102353	587059.275344	60983.0	4.0	60.0	15.0	4.0	18.0	103.0	51.567281	-0.271057	4
9	Ealing	1.087245e+06	661041.885124	541934.840918	59413.0	4.0	68.0	16.0	6.0	21.0	117.0	51.513254	-0.304314	4
12	Hackney	9.603845e+05	909784.311727	772467.509688	65138.0	2.0	58.0	16.0	3.0	32.0	113.0	51.573445	-0.072438	4
15	Harrow	1.052568e+06	570707.666496	455674.838746	31820.0	1.0	42.0	13.0	4.0	13.0	75.0	51.580559	-0.341995	4
21	Kingston upon Thames	9.722135e+05	639571.847528	472129.336597	25390.0	1.0	36.0	11.0	3.0	12.0	64.0	51.412330	-0.300689	4
22	Lambeth	1.120551e+06	895116.223354	698312.693072	70894.0	5.0	60.0	19.0	5.0	8.0	99.0	51.457148	-0.123068	4

Appendix

Data Source

DATA	WEBPAGE LINK
UK House Price Index	https://www.gov.uk/government/publications/about-the-uk-house-price-index
Data Dictionary	https://www.gov.uk/government/publications/about-the-uk-house-price-index/about-the-uk-house-price-index
Recorded Crime: Geographic Breakdown	https://data.london.gov.uk/dataset/recorded_crime_summary
Schools and Pupils by Type of School, Borough	https://data.london.gov.uk/dataset/schools-and-pupils-type-school-borough
List of London Boroughs	https://en.wikipedia.org/wiki/List_of_London_boroughs

Python Notebook

Github Repo	https://github.com/linyang1987/Coursera_Capstone/blob/master/Data%20Science%20Final%20Project.ipynb