THE BATTLE OF NEIGHBORHOODS

SEGMENTATION AND CLUSTERING NEIGHBORHOODS IN LONDON

INTRODUCTION

- Optimum location for new real estate investment
- **Business problem:**
- How to find the right location for an investment, if prices even in nearby locations can be significantly different from each other. Is the price of a property also shaped by the level of crime in a given region.
- Success criteria:
- Good recommendation of Borough/Neighborhood for investment in real estate

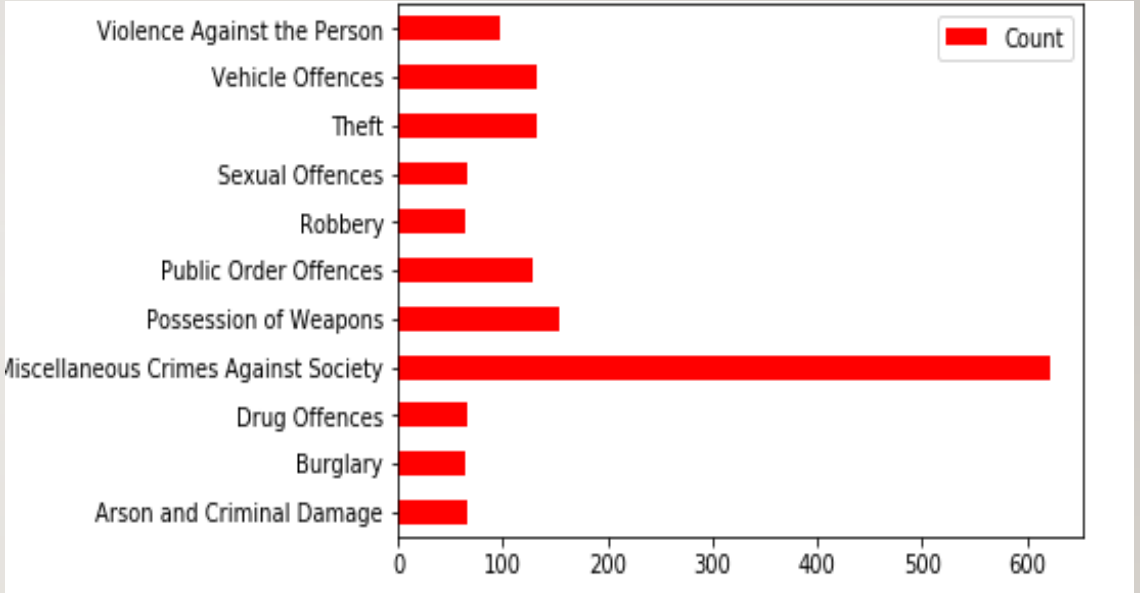
LONDON - FACTS

London is one of the cities in Europe. The number of inhabitants exceeded 8 million people at 1572 km². London is the second fastest center of the world. Several hundred banks and large insurance and investment companies have their headquarters there. It is also a huge media center. London attracts about 30 million tourists a year.

DATA DESCRIPTION

Data 1 – London crime data dived on locations.

1. *MPS Borough Level Crime (most recent 24 months).csv* file.
2. This dataframe contains data about kind of crime, borough name and the data in columns from April of 2017 to march of 2019.



DATA DESCRIPTION

Data 2 – London boroughs Geographical Coordinates data.

WikiPedia
https://en.wikipedia.org/wiki/List_of_London_boroughs

	Borough	Latitude	Longitude
0	BARKING AND DAGENHAM	51.5607	0.1557
1	BARNET	51.6252	-0.1517
2	BEXLEY	51.4549	0.1505
3	BRENT	51.5588	-0.2817
4	BROMLEY	51.4039	0.0198

DATA DESCRIPTION

Data 3 – London real estate prices dived on locations.
pp-monthly-update-new-version.csv obtained from webpage

https://www.gov.uk/guidance/about-the-price-paid-data

	Tui	Price	DateOfTransfer	Postcode	PropertyType	New	Duration	PAON	SAON	Street	Locality	TownCity
0	{85866A65-5D4D-143F-E053-6B04A8C06A15}	83000	2002-02-01 00:00	CF72 9JA	D	N	F	1A	NaN	GELLI ROAD	LLANHARRY	PONTYCLUN
1	{85866A64-68E1-143F-E053-6B04A8C06A15}	80000	2002-04-11 00:00	SL1 5AE	F	N	L	2	NaN	OAKFIELD AVENUE	NaN	SLOUGH
2	{85866A64-6936-143F-E053-6B04A8C06A15}	270000	2002-08-09 00:00	RG31 5DB	D	N	L	YORK LODGE	44	PEGASUS COURT	TILEHURST	READING
3	{85866A64-7718-143F-E053-6B04A8C06A15}	250000	2002-03-15 00:00	TS14 8PR	T	Y	F	STABLE HOUSE, 3	NaN	PEASE COURT	NaN	GUISBOROUGH
4	{85866A64-9067-143F-E053-6B04A8C06A15}	72500	2002-05-31 00:00	EX10 9ES	F	N	L	CHEGWORTH	FLAT 1	ARCOT ROAD	NaN	SIDMOUTH

DATA DESCRIPTION

Data 4 – Londom Bromley boroughs.

https://en.wikipedia.org/wiki/List_of_districts_in_the_London_Borough_of_Bromley

foursquare API to get venues for all neighbors

	Borough	Latitude	Longitude
0	ANERLEY	51.412848	-0.065301
1	APERFIELD	51.316629	0.032717
2	BECKENHAM	51.407094	-0.030318
3	BICKLEY	51.401740	0.043712
4	BIGGIN HILL	51.320037	0.009479

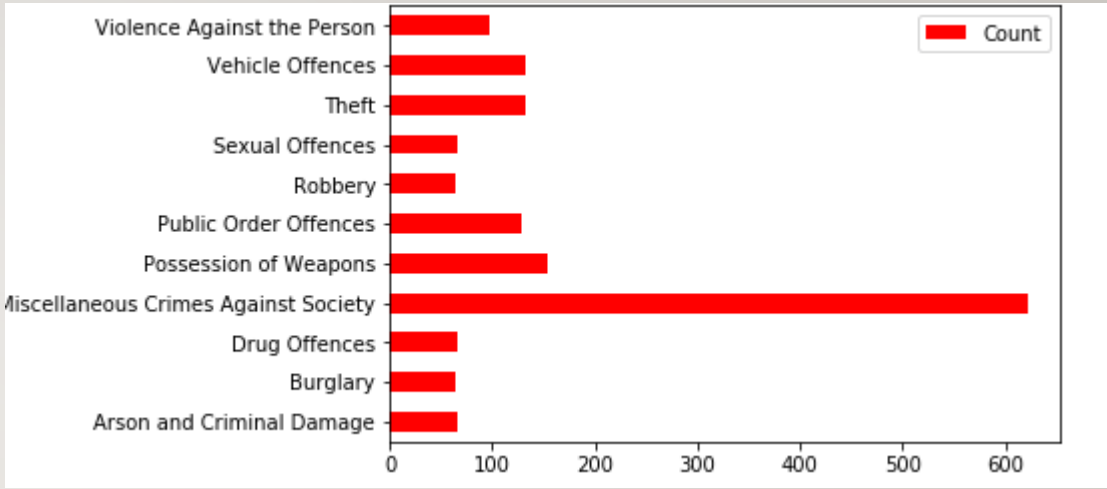
	Location	Location Latitude	Location Longitude	Venue	Venue id	Venue Latitude	Venue Longitude	Venue Category
0	ANERLEY	51.412848	-0.065301	The Douglas fir	56744ee738fa360643ac2d5a	51.414766	-0.070820	Pub
1	ANERLEY	51.412848	-0.065301	BP	508d079be4b055a758969168	51.413183	-0.068181	Gas Station
2	ANERLEY	51.412848	-0.065301	Wickes	59d7690f86f4cc7f7d1d97dc	51.412143	-0.065070	Hardware Store
3	ANERLEY	51.412848	-0.065301	Twang Guitars	4ef4550aa69d3d38d61fec2b	51.415961	-0.064454	Music Store
4	ANERLEY	51.412848	-0.065301	Betts Park	4cbadba60180721e6d5e9c61	51.408755	-0.067278	Park

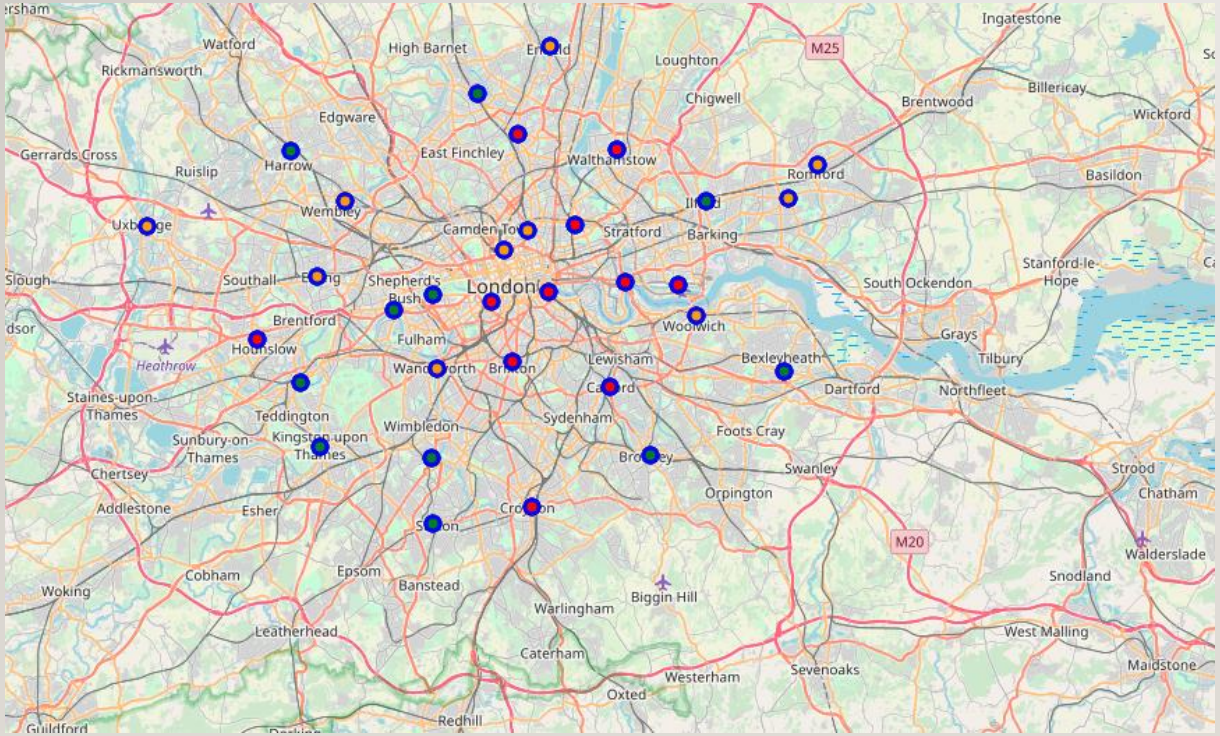
METHODOLOGY

London has many districts.
PART I analyzing the crime data and clustering London district according to crime level.
PART II analyzing the London boroughs according to prices of real assets.
PART III analyzing the veneus in Bromley

METHODOLOGY - I

load and explore data from *MPS Borough Level Crime (most recent 24 months).csv* file.
Use matplotlib to visualize the crime types
Download geolocation data from WikiPedia webpage using python BeautifulSoup library and visualise data.

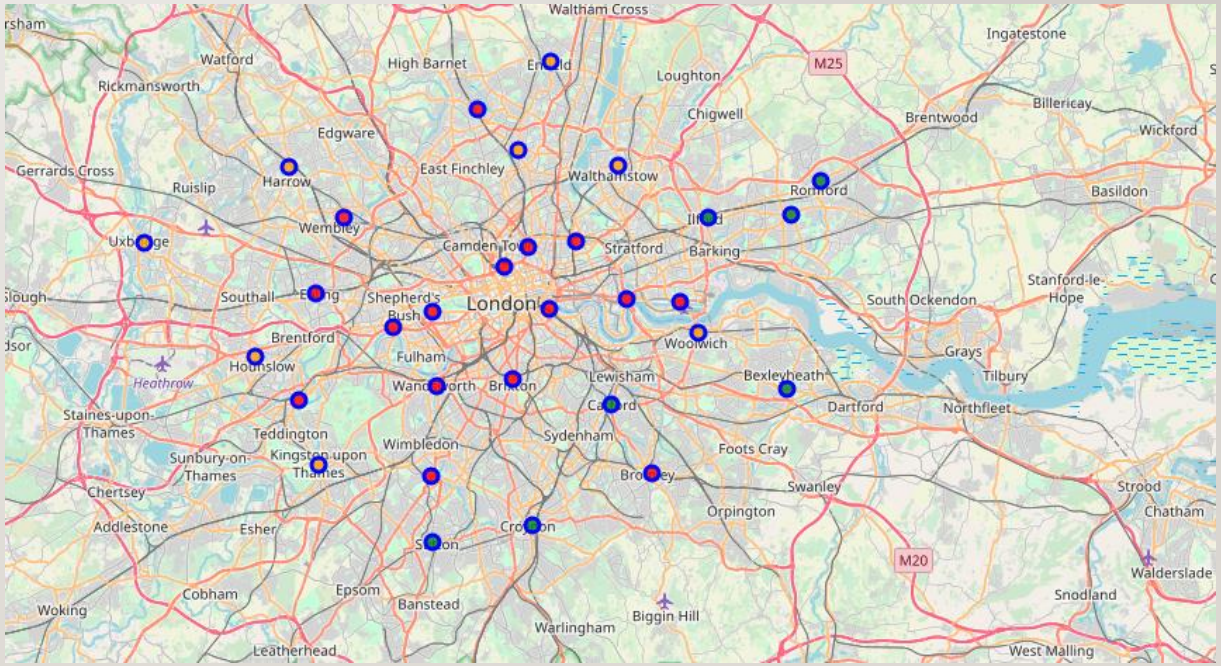




METHODOLOGY – II

Upload data a bout real estate prices from *pp-monthly-update-new-version.csv* file.
Join the data of real assets price with Geographical Coordinates in dataset.
Visualize the data using folium library

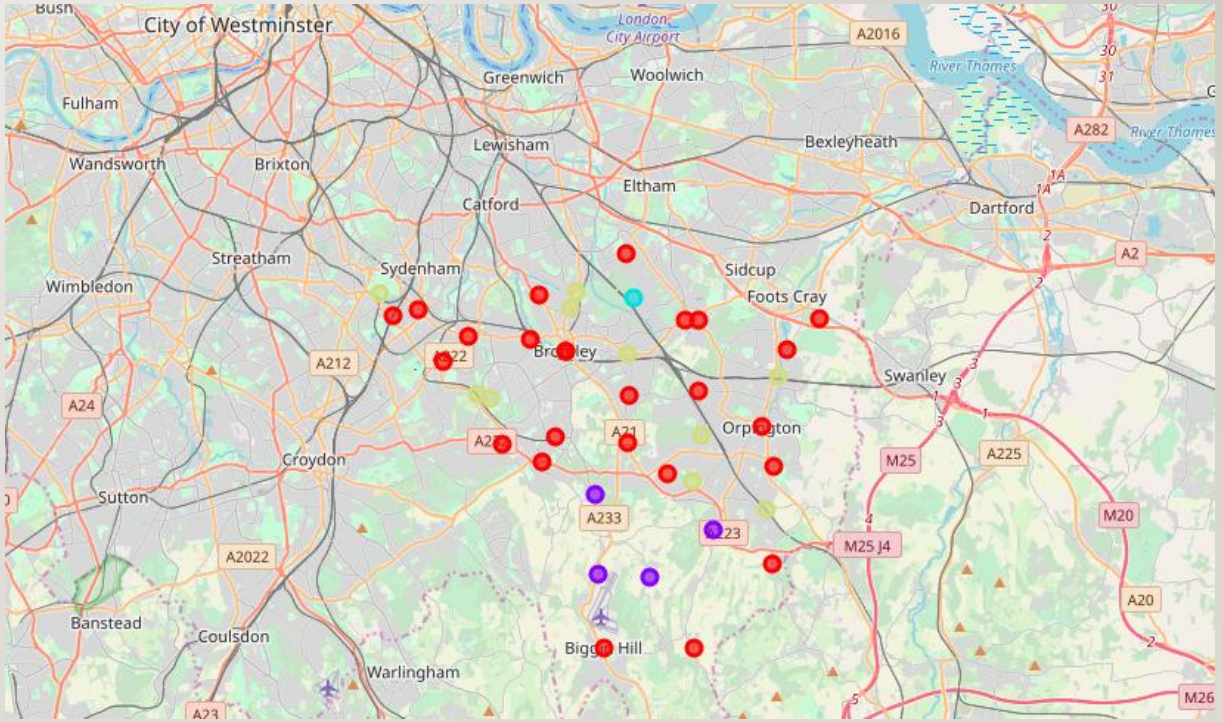
	District	Price	sum	Latitude	Longitude
0	BARKING AND DAGENHAM	344924.678112	23.0	51.5607	0.1557
1	BARNET	868022.700229	12.0	51.6252	-0.1517
2	BEXLEY	388745.295820	12.0	51.4549	0.1505
3	BRENT	699025.365462	18.0	51.5588	-0.2817
4	BROMLEY	762518.715859	15.0	51.4039	0.0198



METHODOLOGY – III

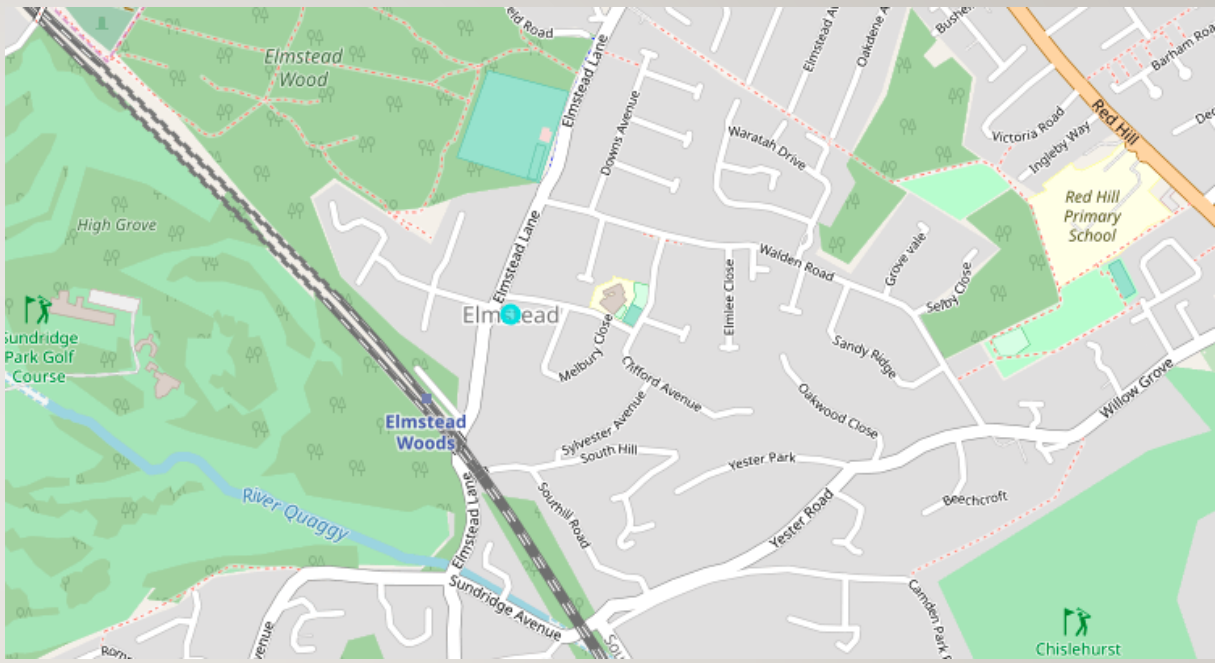
Download data from WikiPedia webpage using python BeautifulSoup library.
Transform the data from webpage and put it into pandas dataframe.
Use geolocator agent to obtain Latitude and Longitude of each borough.
Use foursquare API to get venues for all neighbors
Cluster Brumley locations using Kmeans.

	Borough	Latitude	Longitude
0	ANERLEY	51.412848	-0.065301
1	APERFIELD	51.316629	0.032717
2	BECKENHAM	51.407094	-0.030318
3	BICKLEY	51.401740	0.043712
4	BIGGIN HILL	51.320037	0.009479



RESULT AND DISCUTION

Our analysis showed that there are many places in London with low or medium crime levels. Average real estate prices in these places are higher than in places with high crime levels. Using the KMeans method, we found a district of London Bromley, which stood out from other places in London. We have made a thorough analysis using the Foursquare API for the Bromley location. On its basis, the location of **ELMSTEAD** was selected, which may be a good place for future investments. Available data allowed us to analyze the number and type of crimes committed in a given area. However, the available data is not specific enough for a deeper analysis of the offenses committed. It would be interesting to check, for example, which ethnic groups are committed by particular types of crime. Are crimes committed by residents of a given district, or visitors. On the other hand, due to the protection of personal data, such analyzes can be performed only by authorized persons. When viewing the visualization of the types of crimes committed, it can be seen that the largest number of committed crimes is classified as Miscellaneous Crimes Against Society. It is not known what exactly is hidden under this concept. The data allowed to clearly confirm the correlation between the number of committed crimes and the price of real estate. However, it would be necessary to check whether these prices result only from low crime levels, or whether there are other factors determining the prices of real estate in a given place. Interesting for investors may be that there are places, such as Bromley, where real estate prices are high and there is a low crime rate. I would suggest to be interested in this place for future investments. In particular, some areas of Bromley may be attractive for new investments due to the lack of, for example, hotels in some areas.



CONCLUSION

Using publicly available data, we could do general analysis about crime in various districts of London. Based on data from real estate purchase and sale transactions, it was possible to check whether there is a correlation between average real estate prices and the level of crime in a given region. Using the foursquare API, we could analyze the most popular venues in the city you are interested in. Unfortunately, data analysis requires searching for data sets in various places, including WikiPedia websites. It is a pity that full data collections are not published on government websites. On the other hand, more detailed data would be necessary for detailed analyzes, and those due to the protection of personal data cannot be public. We discovered place for feature investment, but final decision on optimal location will be made by stakeholders.