The Battle of Neighborhoods – report

1. Introduction and Business Problem.

Problem Background:

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.

It should be remembered that the market is very competitive and has a very important meaning for surrounding cities.

There are more and more different types of crimes in London.

Therefore, ten factors should also be provided.

Problem description:

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. Even if we decide on a location, in which place and what kind of business we should open. Is it supposed to be a restaurant, a bar, or maybe a rental of tennis equipment?

Therefore, the location for the investment must take into account many factors.

Target Audience:

Developers, sellers and buyers of real estate, as well as for companies that want to build new offices in interesting districts.

Success Criteria:

The success criteria of the project will be a good recommendation of Borough/Neighborhood for investment in real estate.

2. Data

To understand and explore I will use the below open data:

1. London DataStore: https://data.london.gov.uk/dataset/recorded_crime_summary

   The file is in csv format. The data includes name of crime, the BoroughName and the number of crimes. The data will be imported to Data Frame.

   Example of data:

	MajorText	MinorText	LookUp_BoroughName	201704	201705	201706	201707	201708	201709	201710	...	201806	20
0	Arson and Criminal Damage	Arson	Barking and Dagenham	13	6	14	2	5	8	7	...	12	
1	Arson and Criminal Damage	Criminal Damage	Barking and Dagenham	139	147	150	143	169	134	132	...	123	
2	Burglary	Burglary - Business and Community	Barking and Dagenham	32	29	19	42	30	25	23	...	33	
3	Burglary	Burglary - Residential	Barking and Dagenham	101	129	71	95	83	81	122	...	77	
4	Drug Offences	Drug Trafficking	Barking and Dagenham	4	4	6	7	1	6	6	...	6	

2. Gov.uk HM Land Registry Price Paid Data: https://www.gov.uk/guidance/about-the-price-paid-data

   The file is in csv format. The data includes among others Price, Property Type, Street, Locality, District.

   Example of 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

3. Foursquare for developers access to venue data: https://foursquare.com/

   Foursquare data provides information about venues in the Borough. This data will be joined with data from above data.

	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

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

   To obtain the list of London Boroughs with geodata, and put it to the data frame.

   Example of data:

	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

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

   To obtain the list of Boroughs one of the London District Bromley for detailed analysis.

	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

3. Metodology

Business understanding

In this project we will direct our efforts on detecting areas of London that have low level of crime and high value of real estates. We also in detail will analyses one of the London district Bromley. The reason is, that Bromley has low level of crimes compare to the neighbors and probably high level of real estate value. We will check it in further analysis.

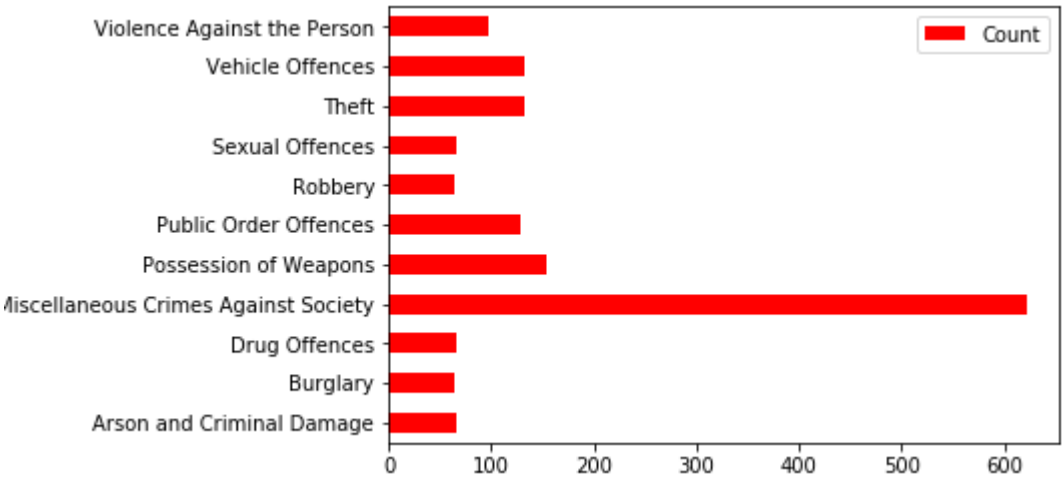
Analytic Approach:

London has many districts. In this project firstly we will analyze the crime data and clustering London district according to crime level. Secondly we will analyze the London boroughs according to prices of real assets. We cluster the boroughs according to the average price in the boroughs.

Exploratory Data Analysis:

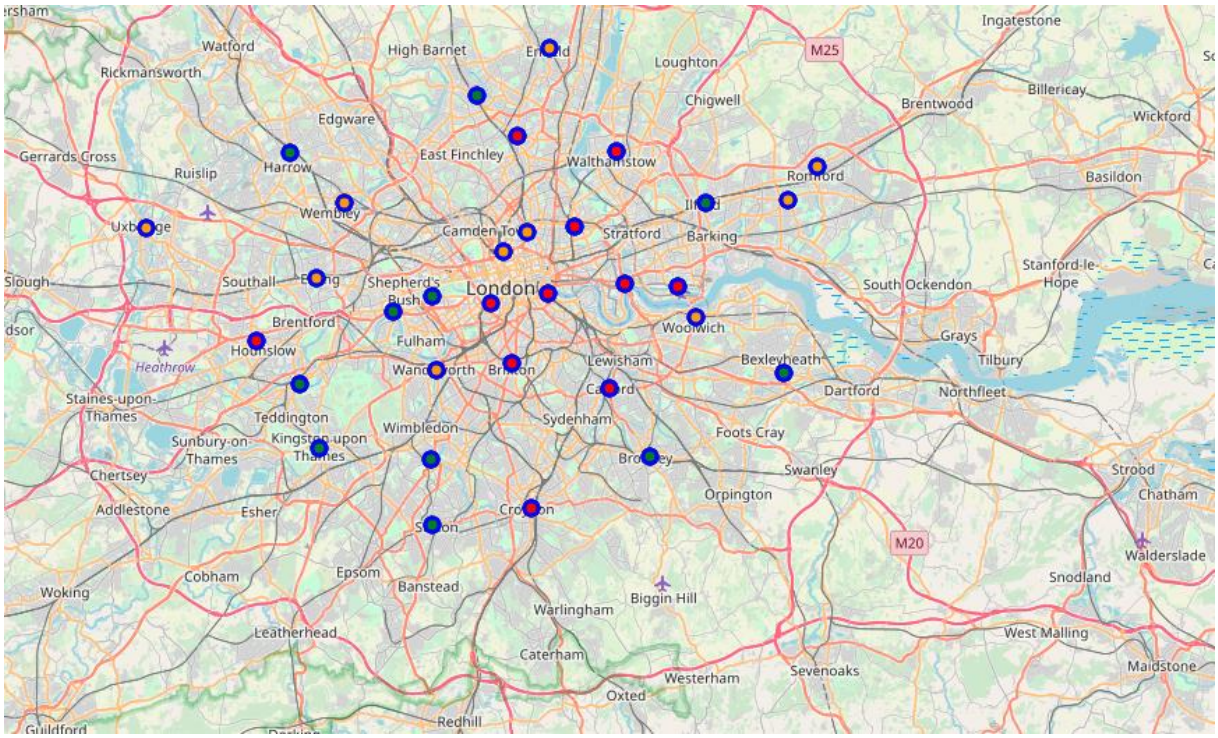
Data 1 – London crime data dived on locations.

1. In this we load and explore data from *MPS Borough Level Crime (most recent 24 months).csv* file.
2. Transform the data and download to pandas dataframe.
3. This dataframe contains data about kind of crime, borough name and the data in columns from April of 2017 to march of 2019.
4. Because we analyze the most recent data we drop all columns except march 2019.
5. We use matplotlib to visualize the crime types.



Data 2 – London boroughs Geographical Coordinates data.

1. In this we download data from WikiPedia webpage using python BeautifulSoup library.
2. We transform the data from webpage and put it into pandas dataframe.
3. We join together the crime data and geographical coordinates data based on borough name.
4. We cluster the data based on crime level of each borough.



Data 3 – London real estate prices dived on locations.

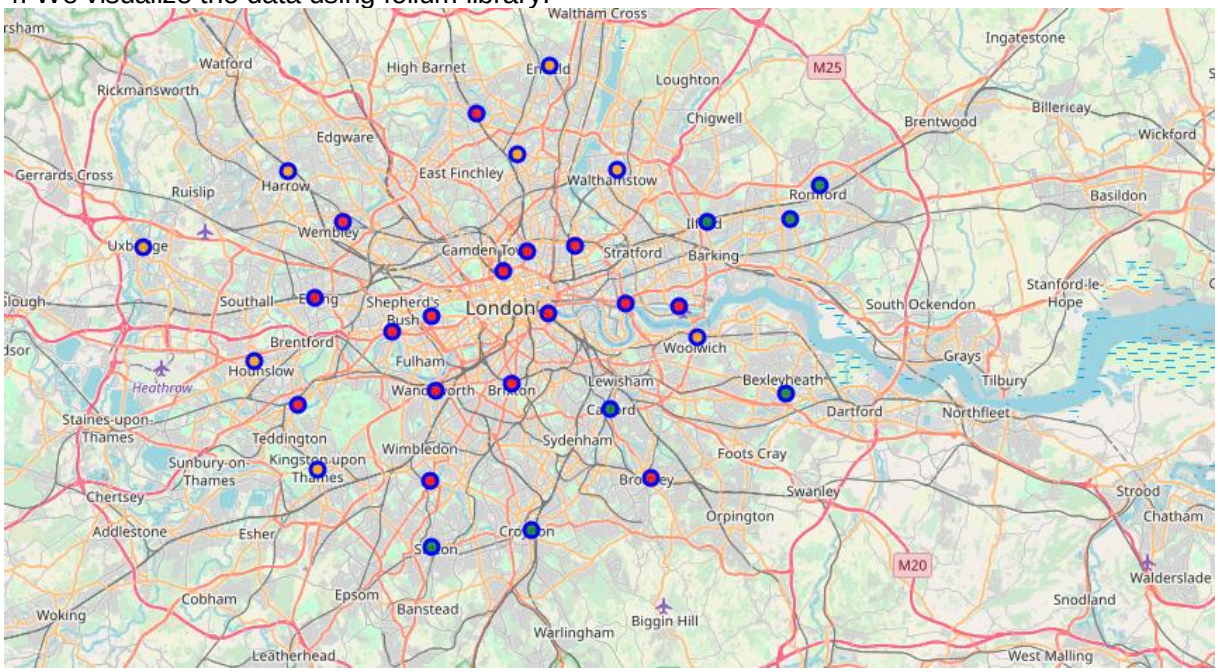
1. In this we upload data from *pp-monthly-update-new-version.csv* file.
2. We drop all rows except the data from London boroughs.

	Tui	Price	DateOfTransfer	Postcode	PropertyType	New	Duration	PAON	SAON	Street	Locality	TownCity
17	{85866A64-EC32-143F-E053-6B04A8C06A15}	115000	2002-10-29 00:00	SE4 2SS	T	N	F	83	NaN	FOXBERRY ROAD	NaN	LONDON
27	{3412C1B2-811B-41D0-A69C-539C980D90FD}	110000	2002-08-16 00:00	E9 5RH	T	N	L	3B	NaN	MABLEY STREET	NaN	LONDON
42	{D1A0C383-5B35-4D8D-9CE1-F183C804FC49}	117000	2002-12-18 00:00	E11 1PP	F	N	L	8A	NaN	HOLLYBUSH HILL	NaN	LONDON
53	{57287FDF-7706-460A-BEFD-8831A837D118}	223000	2002-05-28 00:00	TW1 4SJ	F	N	L	97	GROUND FLOOR FLAT	STRAWBERRY VALE	NaN	TWICKENHAM
55	{175BEBD5-1FD8-494F-834C-B25D89A9A82F}	107000	2002-02-12 00:00	SE14 5NS	F	N	L	39C	BASEMENT FLAT	OMMANEY ROAD	NEW CROSS	LONDON

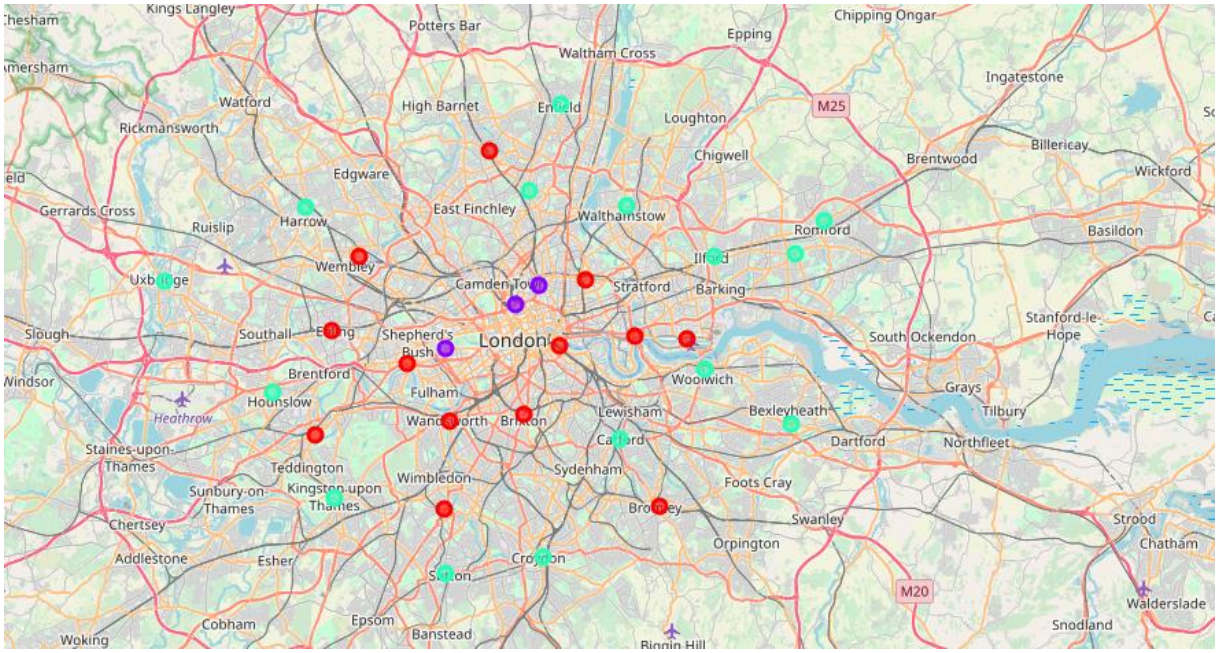
3. We join the data of real assets price with Data 2 to get the Geographical Coordinates in dataset.

	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

4. We visualize the data using folium library.



5. Using the all above data we cluster London locations using KMeans function



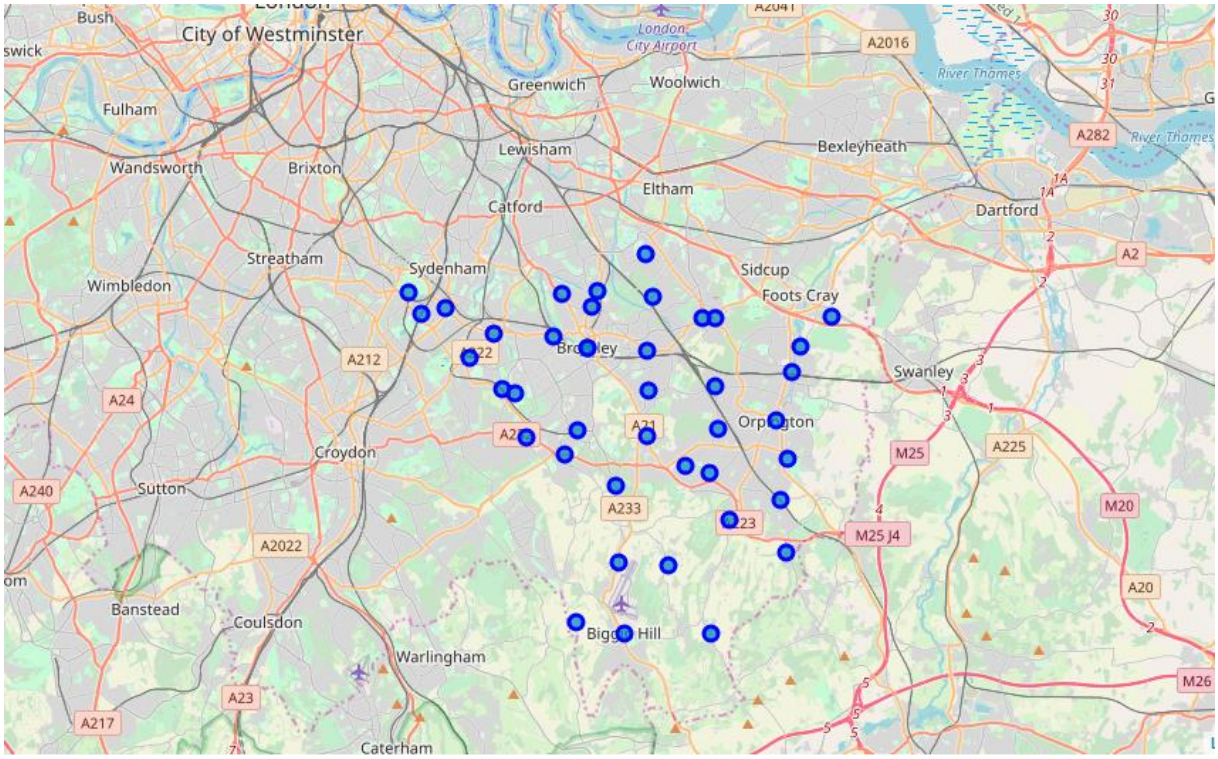
Data 4 – Londom Bromley boroughs.

In our project I decided to analyze in detail one of the district of London Bromley. For further analysis, I need data about Bromley boroughs with geolocation. I found the WikiPedia website with a list of boroughs. Having the list of boroughs, I will use the geolocation API to get the geolocation of each boroughs.

1. In this we download data from WikiPedia webpage using python BeautifulSoup library.
2. We transform the data from webpage and put it into pandas dataframe.
3. We use geolocator agent to obtain Latitude and Longitude of each borough.

	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

4. We visualize the data using folium map.



5. We use foursquare API to get venues for all neighbors.

	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

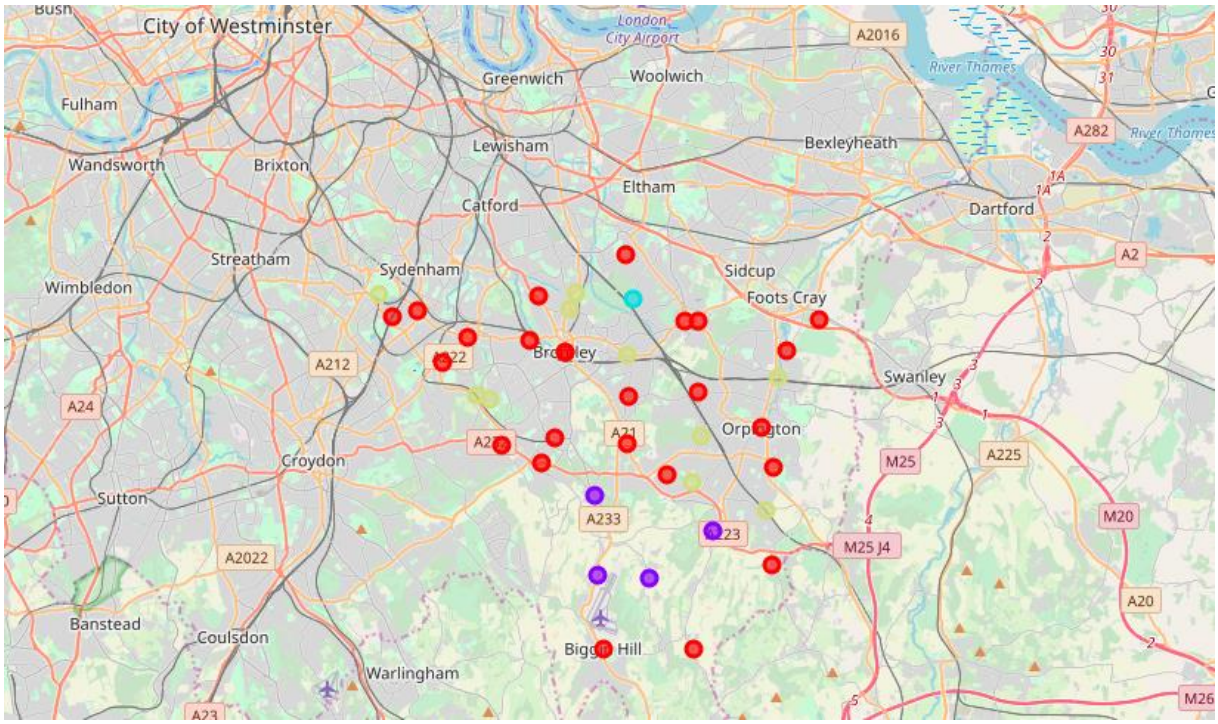
6. We grouping by location and average frequency of occurrence of each venue category

	Location	Airport Terminal	American Restaurant	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Baby Store	Bakery	Bar	Bed & Breakfast	...	Sports Club	Supermarket	Sushi Restaurant
0	ANERLEY	0.000000	0.000000	0.0000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
1	APERFIELD	0.333333	0.000000	0.0000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
2	BECKENHAM	0.000000	0.000000	0.0000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	...	0.000000	0.066667	0.000000
3	BICKLEY	0.000000	0.000000	0.0000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
4	BROMLEY	0.000000	0.000000	0.0000	0.033333	0.000000	0.0000	0.000000	0.033333	0.000000	...	0.000000	0.000000	0.000000
5	BROMLEY COMMON	0.000000	0.000000	0.0000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
6	CHELSFIELD	0.000000	0.000000	0.0000	0.000000	0.000000	0.0000	0.000000	0.250000	0.000000	...	0.000000	0.000000	0.000000
7	CHISLEHURST	0.000000	0.000000	0.0000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
8	CHISLEHURST COMMON	0.000000	0.000000	0.0000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
9	CONEY HALL	0.000000	0.000000	0.0000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
10	CROFTON	0.000000	0.000000	0.0000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
11	CRYSTAL PALACE	0.000000	0.000000	0.0000	0.000000	0.045455	0.0000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
12	CUDHAM	0.000000	0.000000	0.0000	0.000000	0.000000	0.0000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000

7. We produce Rank of the top most common venues

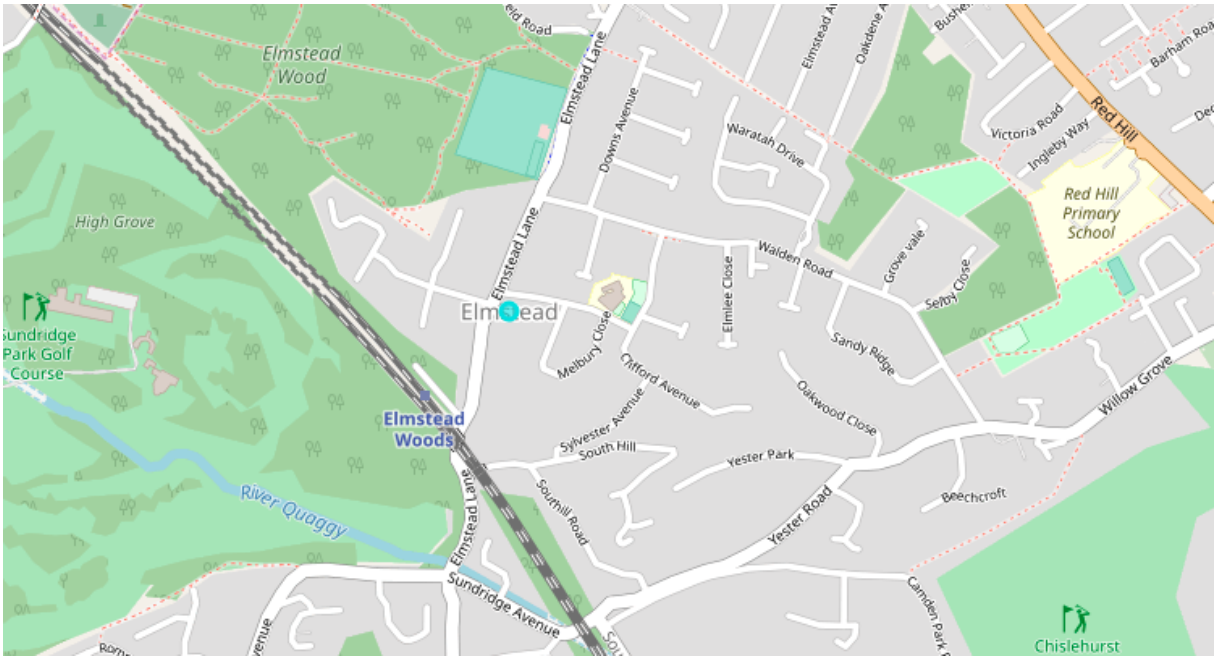
	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	ANERLEY	Grocery Store	Gas Station	Park	Hardware Store	Music Store	Pub	Wine Shop	Donut Shop	Cricket Ground	Dance Studio
1	APERFIELD	Airport Terminal	Pub	Music Venue	Café	Grocery Store	Athletics & Sports	Farm	Dance Studio	Deli / Bodega	Department Store
2	BECKENHAM	Coffee Shop	Italian Restaurant	Pub	Supermarket	Café	Park	Indian Restaurant	Mediterranean Restaurant	Grocery Store	Portuguese Restaurant
3	BICKLEY	Train Station	Cricket Ground	Home Service	Café	Wine Shop	English Restaurant	Dance Studio	Deli / Bodega	Department Store	Diner
4	BROMLEY	Coffee Shop	Clothing Store	Pub	Gym / Fitness Center	Pizza Place	Chocolate Shop	Sandwich Place	Department Store	Park	Donut Shop

8. We Cluster Brumley locations using KMeans



4. Results and discussion

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.

5. 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.