Capstone Project

The Battle of the Neighborhood

Final Presentation

Introduction

- When people, specially from a foreign country, buying houses in a new neighborhood, few things are considered
- Main thing is crime.
- So the how can we find the best place to live in London
- Mainly using crime, and then can we cluster them with other facilities

Data

- We need a few types of data

1) Neighborhoods in London
2) Crimes in those neighborhoods
3) Location data
4) Other facilities around them

How the data was collected

- London boroughs data using Wikipedia

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

- Crime data using Kaggle database

https://www.kaggle.com/jboysen/london-crime

- API geocoding for location data

Data Cleaning

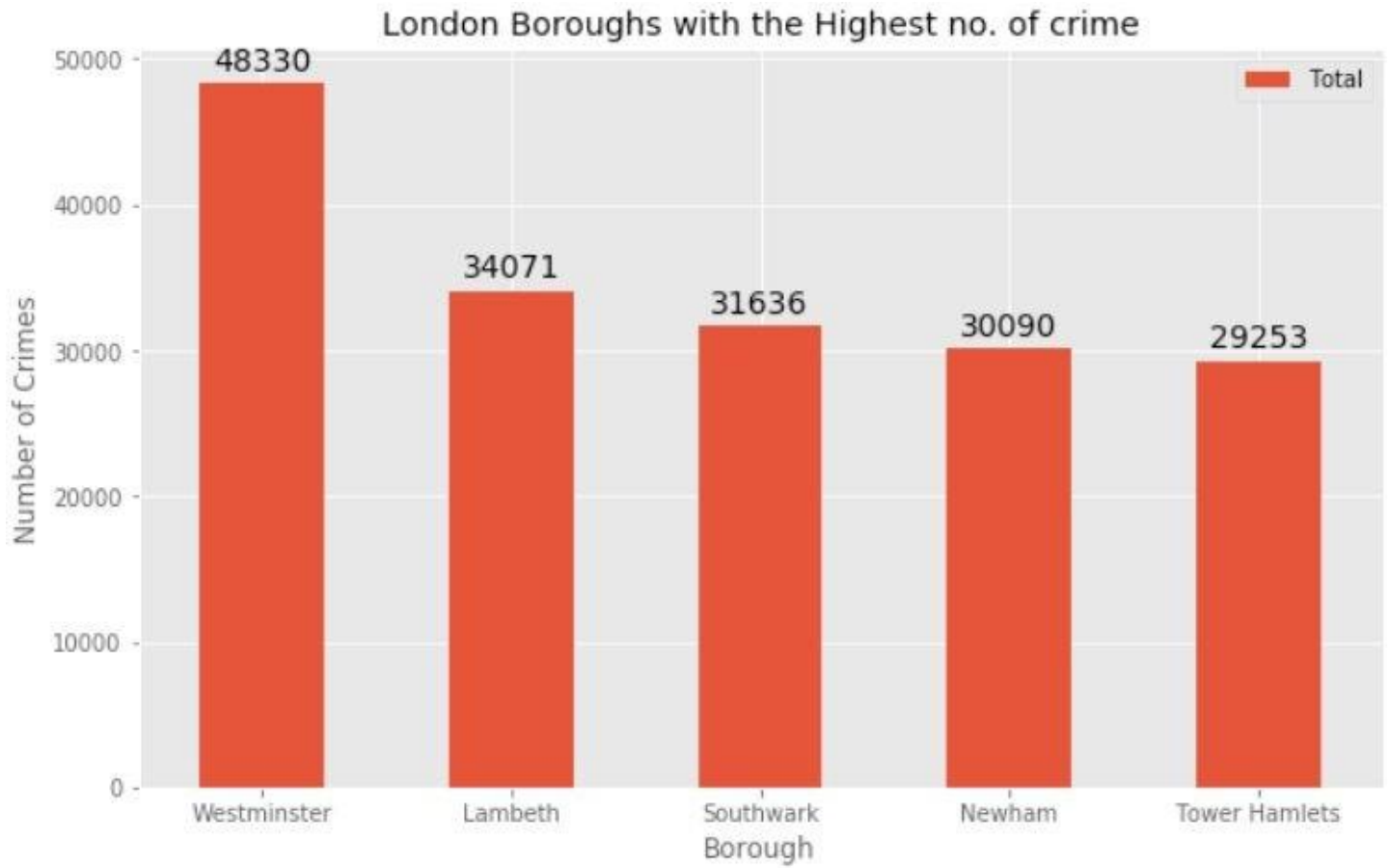
- For the London Crime data, due to the size of the database is high, only the recent 2 years was selected assuming that current behavior will be reflected by those 2
- Wikipedia table was scraped from the page
- Then according to the neighborhood those 2 were merged together.
- Considering the crime, neighborhoods with lowest crimes were selected for clustering.

Methodology

- Exploratory Data Analysis

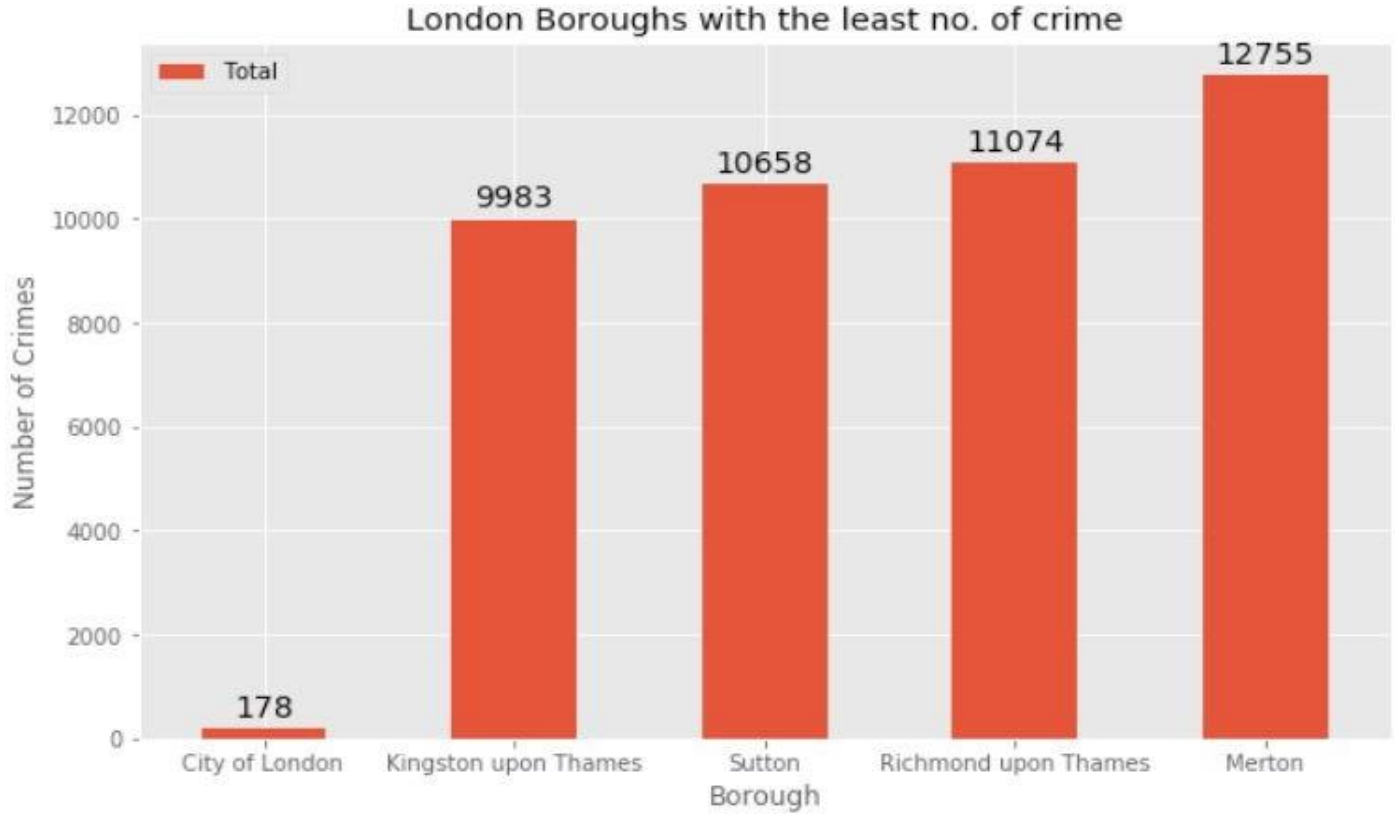
	Burglary	Criminal Damage	Drugs	Other Notifiable Offences	Robbery	Theft and Handling	Violence Against the Person	Total
count	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000	33.000000
mean	2069.242424	1941.545455	1179.212121	479.060606	682.666667	8913.121212	7041.848485	22306.696970
std	737.448644	625.207070	586.406416	223.298698	441.425366	4620.565054	2513.601551	8828.228749
min	2.000000	2.000000	10.000000	6.000000	4.000000	129.000000	25.000000	178.000000
25%	1531.000000	1650.000000	743.000000	378.000000	377.000000	5919.000000	5936.000000	16903.000000
50%	2071.000000	1989.000000	1063.000000	490.000000	599.000000	8925.000000	7409.000000	22730.000000
75%	2631.000000	2351.000000	1617.000000	551.000000	936.000000	10789.000000	8832.000000	27174.000000
max	3402.000000	3219.000000	2738.000000	1305.000000	1822.000000	27520.000000	10834.000000	48330.000000

The count for each of major categories of crimes returns the value 33 and the most common crime is Theft and Handling and the lowest one is Drugs, and other notifiable offenses

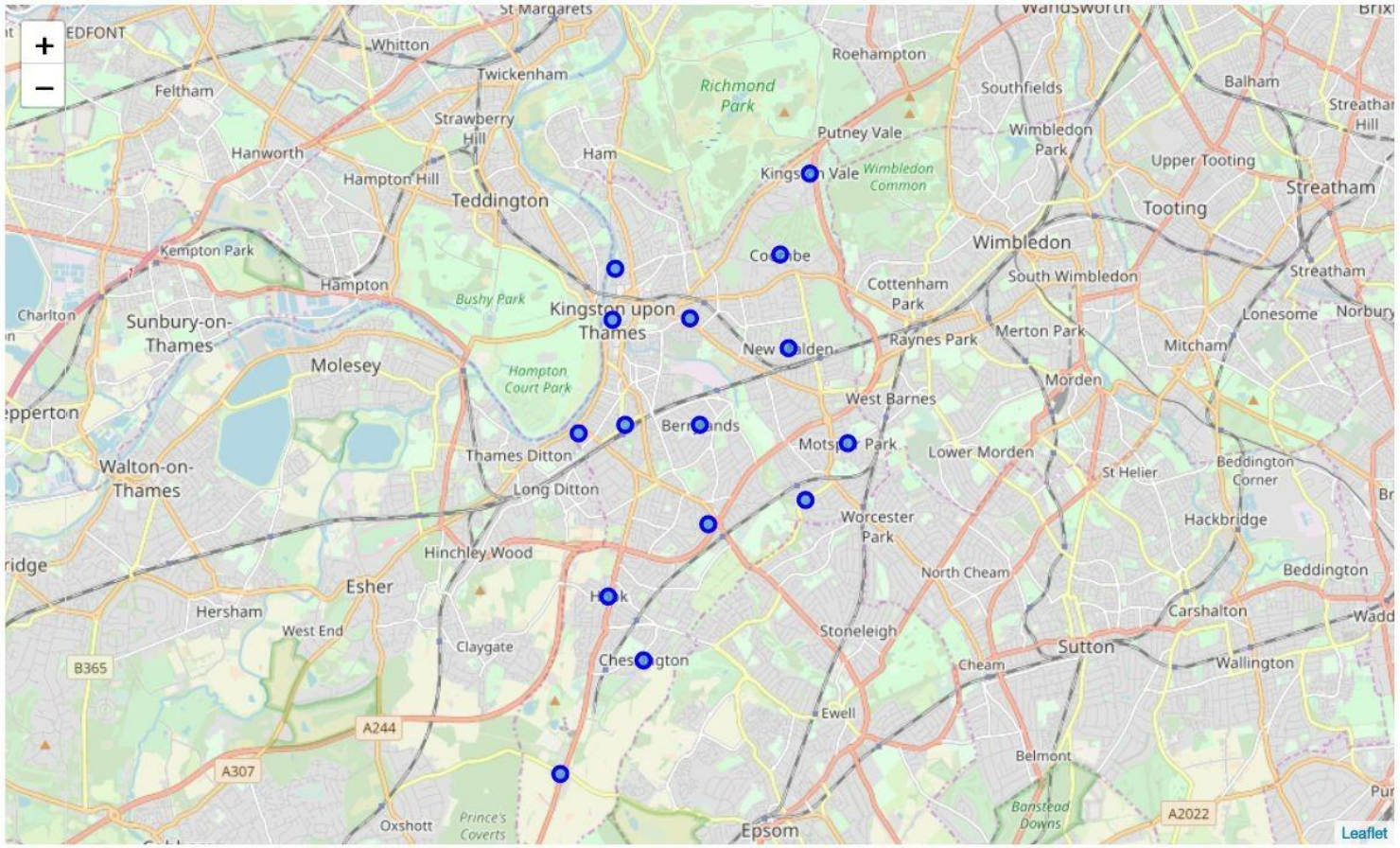


In the year 2016, this graph show the Boroughs with most numbers of crimes.

We consider the highest for the Boroughs which are least peaceful considering the Number of crimes.



- In 2016, this graph show the Boroughs with lowest number of crimes.
- We consider them to be more peaceful and take them for the clustering for the next step



- Theses are the Neighborhoods in Kingston upon Thames, which showed the lowest crime rates
- We use these 15 neighborhood for the best place

Modeling

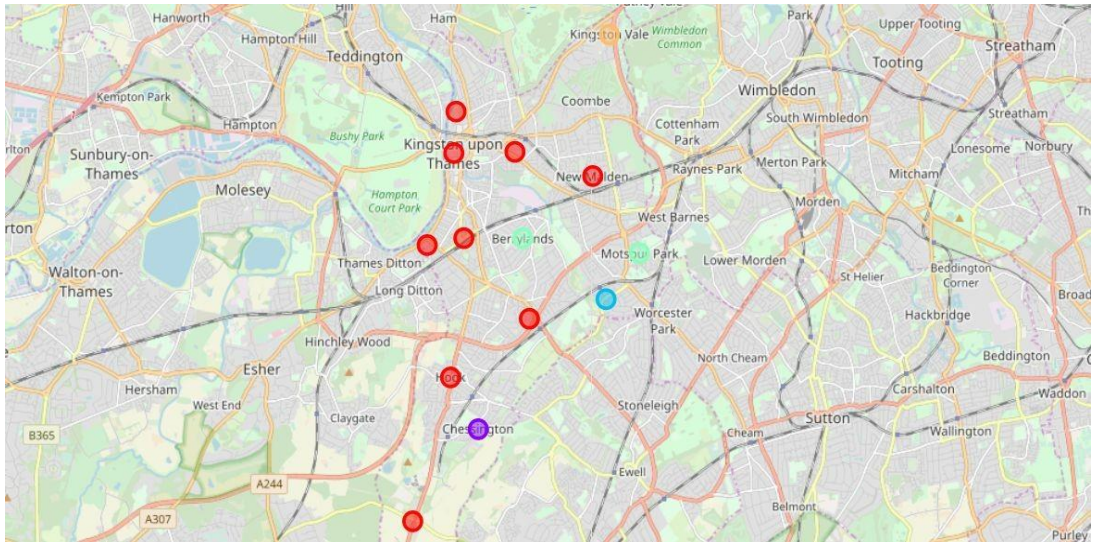
- So we take the Neighborhood data for the Kingston upon Thames Borough

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Berrylands	51.393781	-0.284802	Surbiton Racket & Fitness Club	51.392676	-0.290224	Gym / Fitness Center
1	Berrylands	51.393781	-0.284802	Alexandra Park	51.394230	-0.281206	Park
2	Berrylands	51.393781	-0.284802	K2 Bus Stop	51.392302	-0.281534	Bus Stop
3	Berrylands	51.393781	-0.284802	Cafe Rosa	51.390175	-0.282490	Café
4	Canbury	51.417499	-0.305553	The Boater's Inn	51.418546	-0.305915	Pub

- Then, one hot encoding was done and the venues were grouped
- Goal is to find common venue categories to be clustered.
- Clusters with k=5 was used.

Results

- K value was 5. Using the folium, they can be visualized as,



- Note that this clusters represent the community. That mean, in the Borough of lowest crime rate, we want to find the places according to the customers.
- So above figure represent the clusers.

Let’s consider cluster 1

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	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
1	Canbury	Kingston upon Thames	51.417499	-0.305553	0	Pub	Café	Plaza	Fish & Chips Shop	Supermarket	Spa	Shop & Service	Park
4	Hook	Kingston upon Thames	51.367898	-0.307145	0	Bakery	Convenience Store	Indian Restaurant	Fish & Chips Shop	Wine Shop	Food	Electronics Store	Farmers Market
5	Kingston upon Thames	Kingston upon Thames	51.409627	-0.306262	0	Coffee Shop	Café	Burger Joint	Sushi Restaurant	Pub	Record Shop	Cosmetics Shop	Market
7	Malden Rushett	Kingston upon Thames	51.341052	-0.319076	0	Convenience Store	Pub	Garden Center	Restaurant	Fast Food Restaurant	Discount Store	Dry Cleaner	Electronics Store
9	New Malden	Kingston upon Thames	51.405335	-0.263407	0	Gastropub	Gym	Sushi Restaurant	Supermarket	Korean Restaurant	Indian Restaurant	Fish & Chips Shop	Dry Cleaner
10	Norbiton	Kingston upon Thames	51.409999	-0.287396	0	Indian Restaurant	Pub	Food	Italian Restaurant	Platform	Grocery Store	Farmers Market	Dry Cleaner
12	Seething Wells	Kingston upon Thames	51.392642	-0.314366	0	Indian Restaurant	Coffee Shop	Italian Restaurant	Pub	Café	Wine Shop	Fast Food Restaurant	Chinese Restaurant
13	Surbiton	Kingston upon Thames	51.393756	-0.303310	0	Coffee Shop	Pub	Supermarket	Breakfast Spot	Grocery Store	Gastropub	French Restaurant	Train Station
14	Tolworth	Kingston upon Thames	51.378876	-0.282860	0	Grocery Store	Pharmacy	Furniture / Home Store	Train Station	Pizza Place	Discount Store	Coffee Shop	Bus Stop

- It can be seen that this cluster has some certain venue types more often than other.
- Like that, we can define the results and group them for customers.

Let’s consider cluster 2 and 3 by neighborhood.

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	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
2	Chessington	Kingston upon Thames	51.358336	-0.298622	1	Fast Food Restaurant	Wine Shop	Golf Course	German Restaurant	Gastropub	Garden Center	Furniture / Home Store	Fried Chicken Joint	French Restaurant

This is for cluster 2

	Neighborhood	Borough	Latitude	Longitude	Cluster Labels	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
11	Old Malden	Kingston upon Thames	51.382484	-0.25909	2	Train Station	Pub	Food	Gastropub	Garden Center	Furniture / Home Store	Fried Chicken Joint	French Restaurant	Deli / Bodega

This is for cluster 3

Discussion

- The Aim was to find the most suitable and peaceful venues for the customers who are buying houses in London city.
- We saw that we first narrowed it down to the Boroughs with least crime figures.
- Then, we clustered Kingston upon Thames, which was the least crimes reported, for the clustering.
- In the clustering, we grouped them according to the similarities of the facilities available there.
- For example, cluster 4 has more venues like Parks, Fitness centers while cluster 2 has more pubs, resturants.

Conclusion

- With the development of data science, projects such as this, which was to find the best places for the buyers to buy houses in London can be done using computer processing power instead of doing it manually.
- The last past was clustering. We can clearly see that those process of finding the similar patters of the venues can hardly be done manually. It has to be accomplished with the tools we have here.
- The project show that importance of Data Science.