



# The Battle of Neighborhoods

RELOCATION  
CHALLENGE

BRIAN LAI

JAN 9, 2021

# Problem Statement



Using data to look for a place to relocate within London which has similar characteristics to the current living area.



The family has preference in safety, school quality, travel time commute to work and match with their lifestyle.

# Data Sources

- School quality: Ofsted which is the Office for Standards in Education, Children's Services and Skills. (<https://reports.ofsted.gov.uk/>)
- Crime statistic: Metropolitan Police ([https://data.london.gov.uk/dataset/recorded\\_crime\\_summary](https://data.london.gov.uk/dataset/recorded_crime_summary))
- Population: Office for National Statistics (ONS) Population Estimates, Borough and Ward (<https://data.london.gov.uk/dataset/office-national-statistics-ons-population-estimates-borough>)
- Location and travel time: Google Geocoding API & Direction API
- Lifestyle: Foursquare API

# School Data

- Features required are
  - Local authority area (i.e. borough of the school)
  - Overall effectiveness (Outstanding, Good, Requirement improvement and Inadequate)
  - Learner numbers (i.e. number of students)
- Records outside London are dropped
- REDACTED status are dropped
- Outstanding % calculated as sum of students in outstanding schools divided by total number of students in the same area.

Area	Population	Outstanding	Outstanding_percentage
Barking and Dagenham	58867.00	5414.00	9.20
Barnet	82685.44	24455.48	29.58
Bexley	52221.00	8297.00	15.89
Brent	55590.00	13633.00	24.52
Bromley	79514.00	18497.00	23.26

# Crime and population

- Crime
  - All crime types are sum up in past two years (Dec 2018 - Nov 2020) and group by the Ward code as unique key
- Population
  - Latest available year (2015) is used, directly use the population of all ages and Ward code as unique key
- Two datasets are merged by common key Ward code.
- Crime rate is calculated as number of crime events divided by population of the ward.

	WardCode	TotalCrime	WardName	Area	WardPopulation	CrimeRate
0	E05000026	5691	Abbey	Barking and Dagenham	14179	40.14
1	E05000027	1739	Alibon	Barking and Dagenham	11007	15.80
2	E05000028	2075	Becontree	Barking and Dagenham	13687	15.16
3	E05000029	1964	Chadwell Heath	Barking and Dagenham	10769	18.24
4	E05000030	1623	Eastbrook	Barking and Dagenham	11255	14.42

# Location and travel time

- **Location**
  - Google Geocoding API (Query by ward and borough name)
  - Set 'postal\_town':'London', 'country':'GB' to make the search more accurate
- **Direction**
  - Google Direction API
  - Starting point (Based on location obtained above)
  - Destination - Canary Wharf, London
  - Set transit as rail
  - Departure time as a Monday 8am to simulate time commute to work

	WardCode	WardName	Area	Latitude	Longitude	TravelSecond
0	E05000026	Abbey	Barking and Dagenham	51.543674	0.102531	2858
1	E05000027	Alibon	Barking and Dagenham	51.547616	0.151888	3062
2	E05000028	Becontree	Barking and Dagenham	51.562598	0.122430	3471
3	E05000029	Chadwell Heath	Barking and Dagenham	51.572228	0.141995	3500
4	E05000030	Eastbrook	Barking and Dagenham	51.551440	0.163567	3197

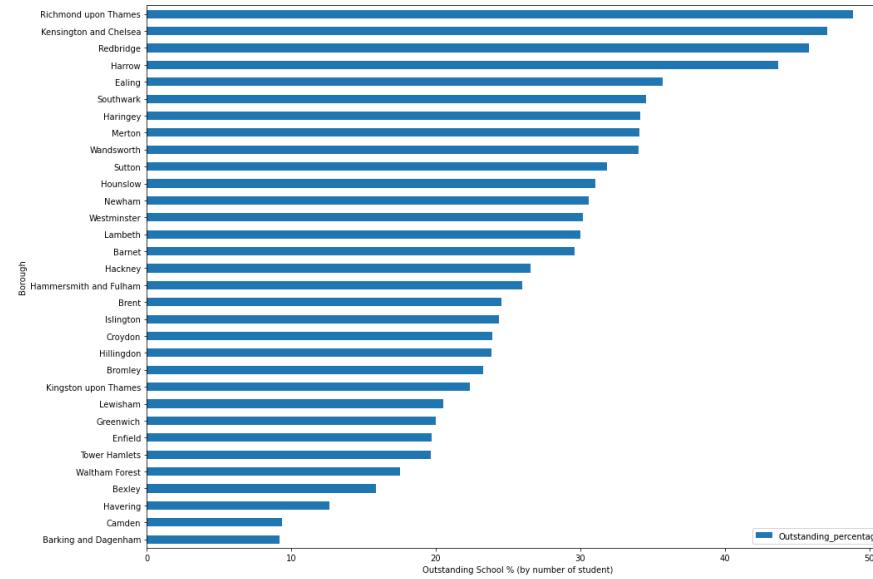
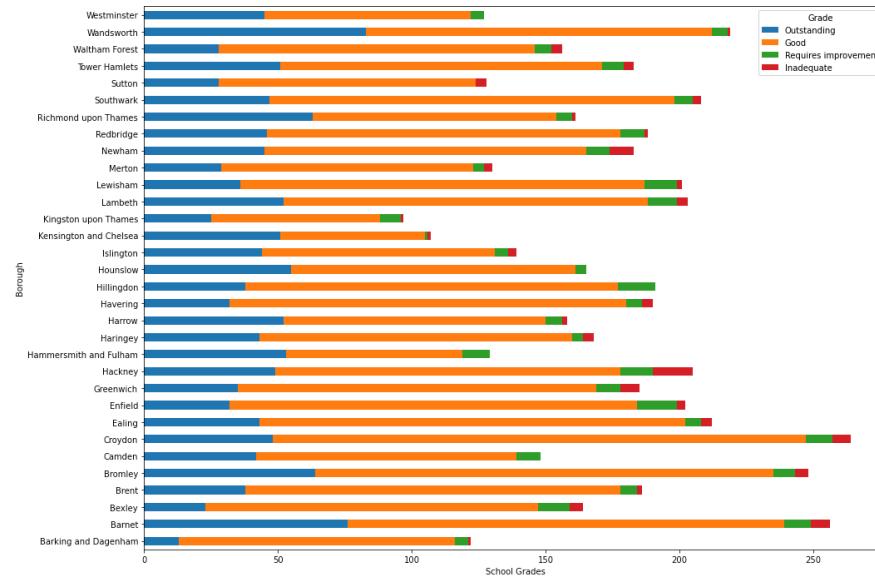


# Lifestyle venues

- **Foursquare API**

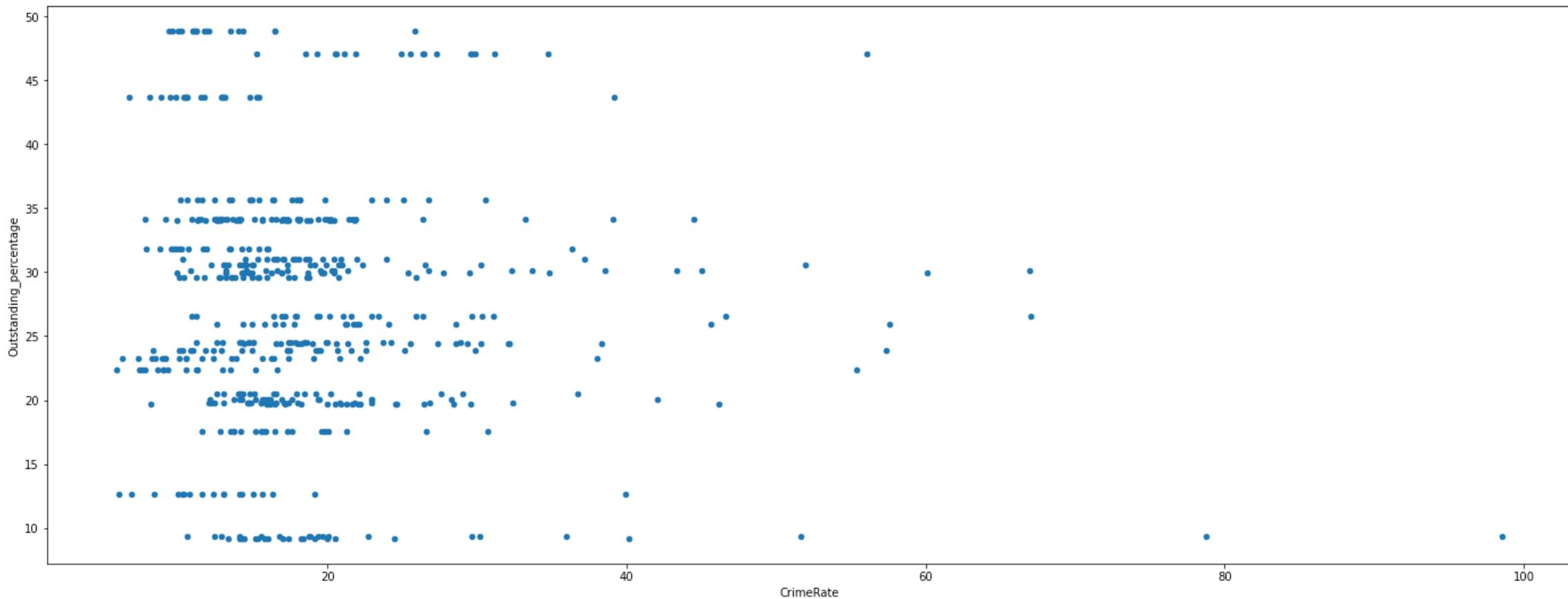
- Search venues within 500 meters of the locations using category ID of Foursquare
  - NIGHTLIFE='4d4b7105d754a06376d81259'
  - RESTAURANT='4bf58dd8d48988d111941735' #JAPANESE
  - COFFEE='4bf58dd8d48988d1e0931735' #COFFEE SHOP

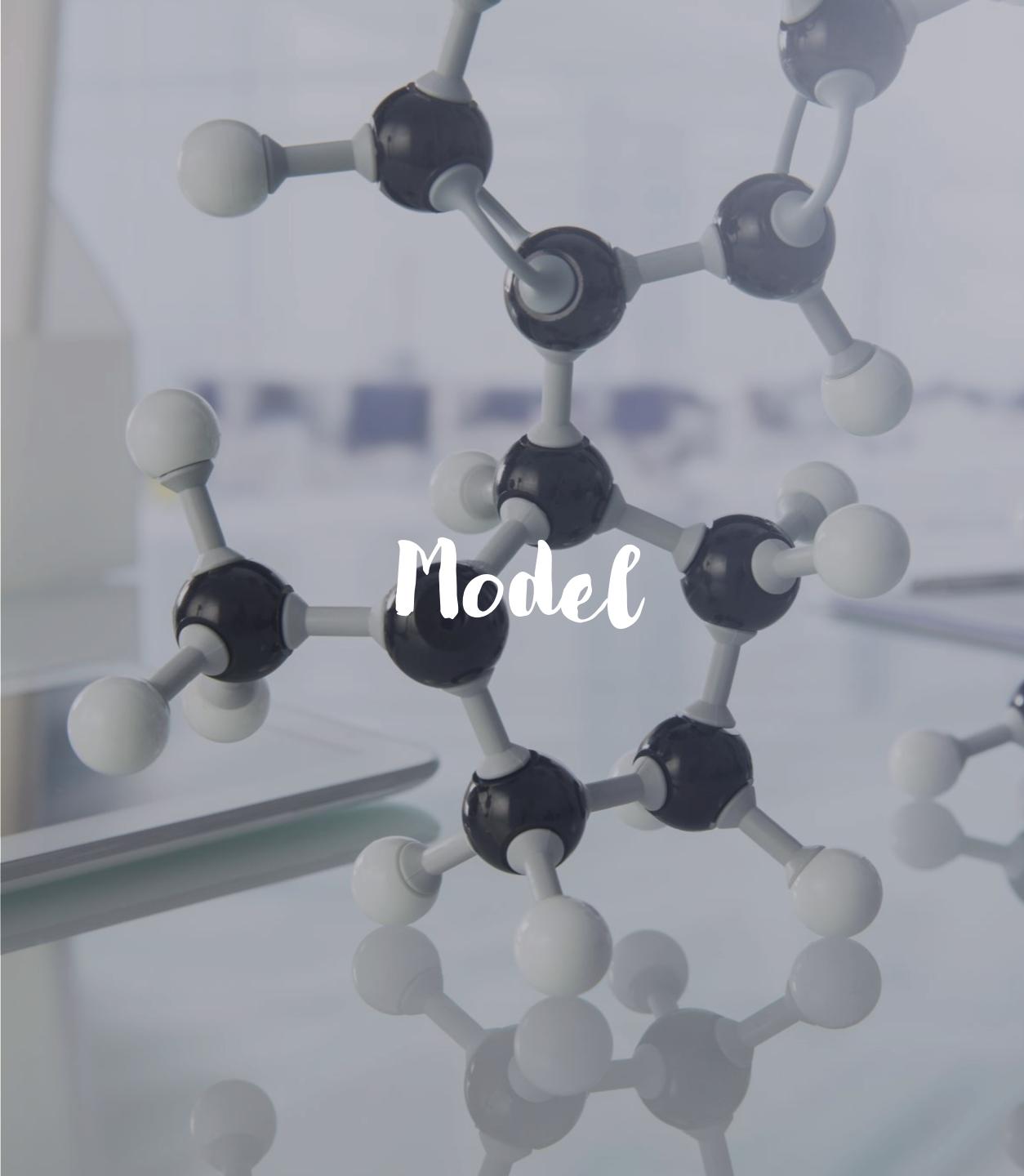
	WardCode	Nightlife	Restaurant	Coffee
0	E05000371	6	2	11
1	E05000380	32	5	22
2	E05000026	0	0	1
3	E05000027	1	0	0
4	E05000028	1	0	2



# School distribution by area

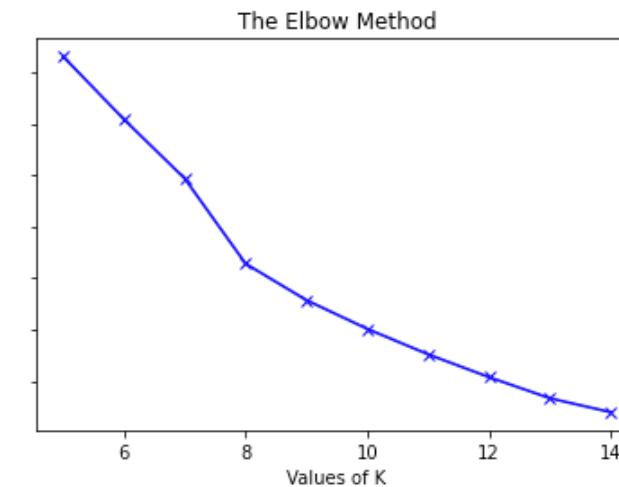
# Relationship Crime vs School





```
array([[ 0.9942901 , -1.90458912, -0.37445219, -0.70020712, -0.33218262,
       -0.54063245],
       [-0.19244876, -1.90458912, -0.20453821, -0.57880204, -0.33218262,
       -0.6450969 ],
       [-0.22365307, -1.90458912,  0.13612266, -0.57880204, -0.33218262,
       -0.43616799],
       ...,
       [ 0.48234441,  2.06600885, -0.89335615,  0.02822332,  0.15906698,
       -0.01831019],
       [ 0.06888733,  2.06600885, -0.25451291, -0.3359919 , -0.33218262,
       -0.54063245],
       [ 0.28390451,  2.06600885, -0.76258902,  1.72789434,  1.6328158 ,
       2.48883664]])
```

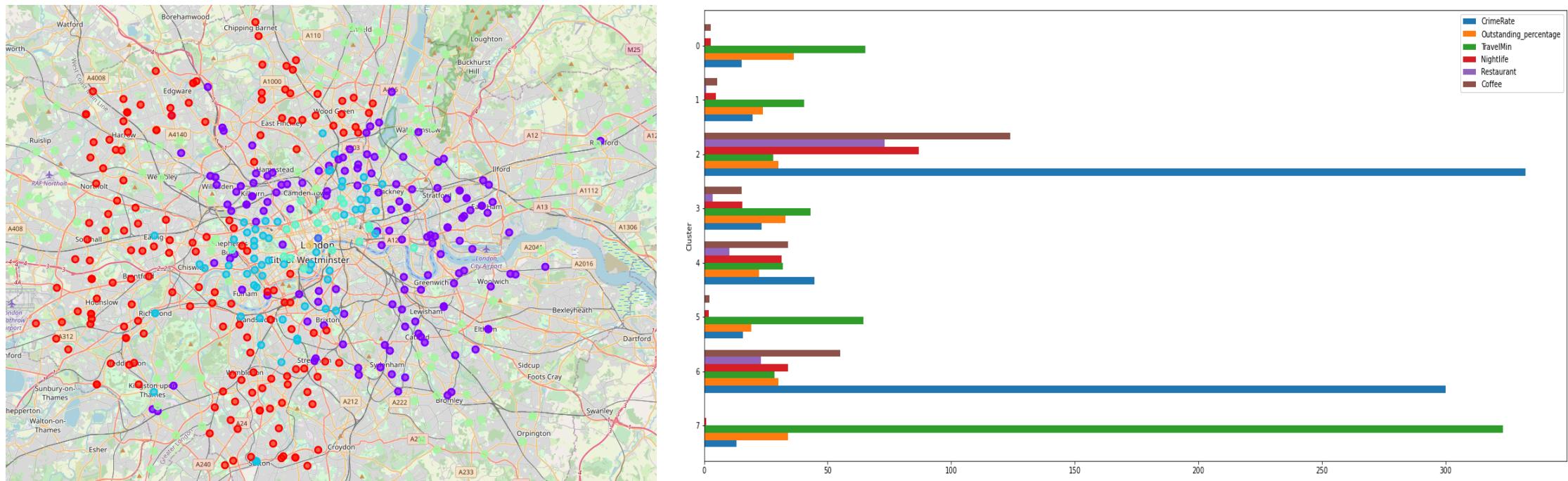
Values are preprocessed with StandardScaler before fit into K-mean model



K - M E A N C L U S T E R I N G  
B Y D I F F E R E N T  
V A L U E O F K

# Result

General speaking, the clustering group are spread by the distance from the central London, with some exception on outliers records with extreme high crime rate / travel time.



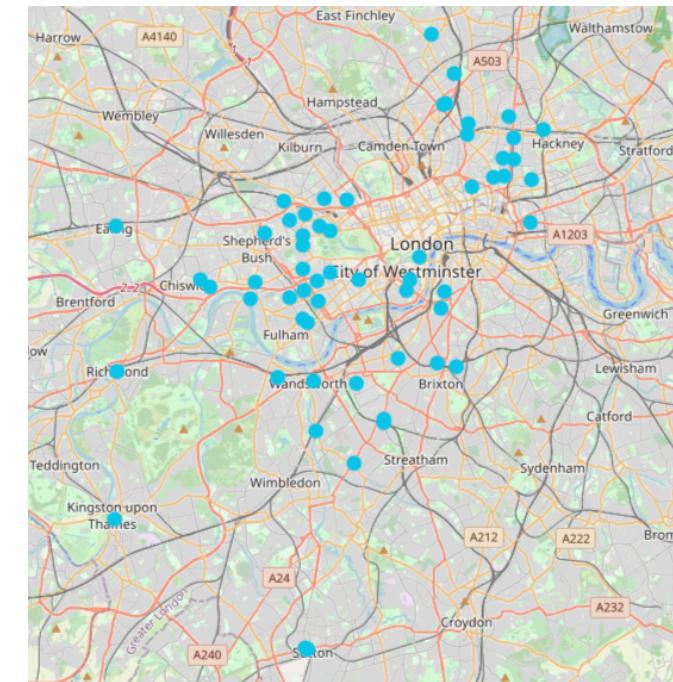
# Discussion

Cluster in the same group as Sutton South

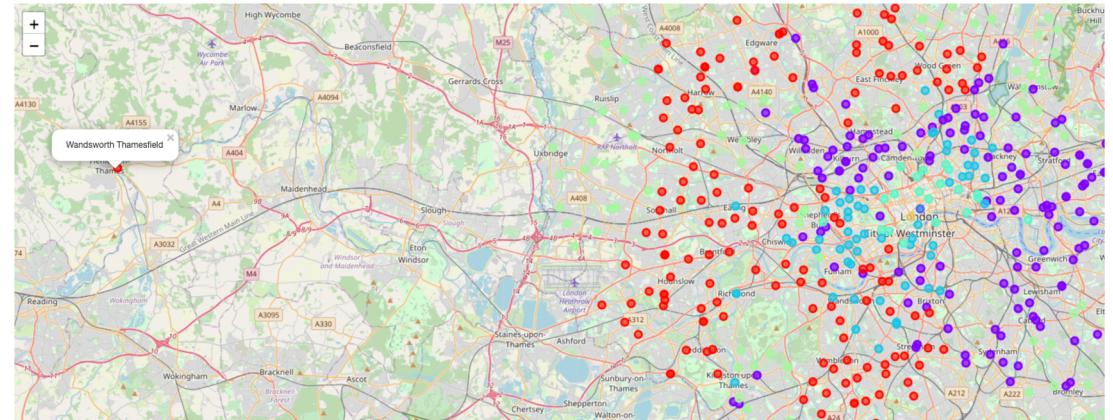
- Relatively low crime rate
- Good outstanding school
- Good travel time
- Some but not a lot lifestyle venues

A few famous outer London areas noted (Richmood, Earling, Kingston upon Thames...etc )

	TotalCrime	WardPopulation	CrimeRate	Population	Outstanding	Outstanding_percentage	Latitude	Longitude	TravelSecond	Nightlife	Restaurant	Coffee	Cluster
count	61.000000	61.000000	61.000000	61.000000	61.000000	61.000000	61.000000	61.000000	61.000000	61.000000	61.000000	61.000000	61.0
mean	2895.278689	12399.131148	23.415410	46933.508197	14597.245902	32.918197	51.494374	-0.164695	2593.196721	15.540984	3.459016	15.327869	3.0
std	1361.904269	2741.010455	9.649576	14508.016882	3013.265486	8.328817	0.047011	0.062649	570.078206	5.858253	1.708545	5.322034	0.0
min	840.000000	5736.000000	9.040000	24788.000000	8598.000000	19.660000	51.361428	-0.304545	1342.000000	6.000000	0.000000	7.000000	3.0
25%	2089.000000	10612.000000	17.260000	41826.000000	11735.000000	26.570000	51.465881	-0.196295	2240.000000	11.000000	2.000000	12.000000	3.0
50%	2639.000000	11919.000000	21.360000	46536.480000	14740.000000	30.160000	51.494197	-0.180248	2536.000000	15.000000	4.000000	15.000000	3.0
75%	3278.000000	14229.000000	27.270000	54511.000000	16440.000000	34.150000	51.523484	-0.114493	2934.000000	19.000000	5.000000	18.000000	3.0
max	8835.000000	20255.000000	67.060000	95269.000000	22216.000000	48.880000	51.581998	-0.057387	4040.000000	28.000000	8.000000	30.000000	3.0



# Further improvement



1. Wrong location return by the Google Geocoding API
  - Better find another data source which provide the location of different Wards directly instead of lookup by name
2. Borough level school data may not be accurate as usually school assigned is near the home
  - Use API or other information to locate nearest school for better estimation of school quality
  - Costly and timely