THE BATTLE OF THE NEIGHBOURHOODS

Applied Data Science Capstone by IBM/Coursera

JUNE 25, 2020

SERCAN KOCAMAZ

Contents

Introduction

London is one of the most popular and multicultural cities in the World. It is diverse and is the financial capital of the UK and Europe. It provides lot of business opportunities and business friendly environment. It is a global hub of business and commerce. The city is a major centre for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theatre, fashion, and the arts in Europe. Therefore, London attracts many people, especially young professionals.

There are many criterions for people who would like to move to London. One of the top concerns is safety when finding a new place. Safety is naturally a really big factor in buying or renting a home. Moreover, it is also important that the area should be lively i.e. close to local amenities. This project aims to find out the top 10 boroughs and explore its neighbourhoods in London.

Data

In order to solve the problem, the following data sources are used in this project.

London Recorded Crime: Geographic Breakdown

Source: London Datastore
https://data.london.gov.uk/dataset/recorded_crime_summary
The main focus is analysing MPS Borough Level Crime (most recent 24 months). Data is taken between 2018-06-01 and 2020-05-31

MajorText	MinorText	LookUp_B	201911	201912	202001	202002	202003	202004
Arson and Crimin	Arson	Barking ar	8	6	4	5	6	2
Arson and Crimin	Criminal Damage	Barking ar	97	121	97	103	108	82
Burglary	Burglary - Business and Cc	Barking ar	30	25	31	17	27	29
Burglary	Burglary - Residential	Barking ar	114	130	116	123	97	56
Drug Offences	Drug Trafficking	Barking ar	12	3	11	3	6	9
Drug Offences	Possession of Drugs	Barking ar	94	79	98	107	102	139
Miscellaneous Cri	Bail Offences	Barking ar	0	0	0	0	0	0
Miscellaneous Cri	Bigamy	Barking ar	0	0	0	0	0	0
Miscellaneous Cri	Dangerous Driving	Barking ar	1	2	2	0	1	0

List of London Boroughs

This is a list of local authority districts within Greater London, including 32 London boroughs and the City of London. The London boroughs were all created on 1 April 1965. Upon creation, twelve were designated Inner London boroughs and the remaining twenty were designated Outer London boroughs. The Office for National Statistics has amended the designations of three boroughs for statistics purposes only.
Source: Wikipedia
https://en.wikipedia.org/wiki/List_of_London_boroughs

Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est)[1]	Co-ordinates	Nr. in map
Barking and Dagenham [note 1]			Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194,352	51.5607°N 0.1557°E	25
Barnet			Barnet London Borough Council	Conservative	North London Business Park, Oakleigh Road South	33.49	369,088	51.6252°N 0.1517°W	31
Bexley			Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236,687	51.4549°N 0.1505°E	23
Brent			Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	317,264	51.5588°N 0.2817°W	12
Bromley			Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	317,899	51.4039°N 0.0198°E	20
Camden	✓		Camden London Borough Council	Labour	Camden Town Hall, Judd Street	8.40	229,719	51.5290°N 0.1255°W	11
Croydon			Croydon London Borough Council	Labour	Bernard Weatherill House, Mint Walk	33.41	372,752	51.3714°N 0.0977°W	19

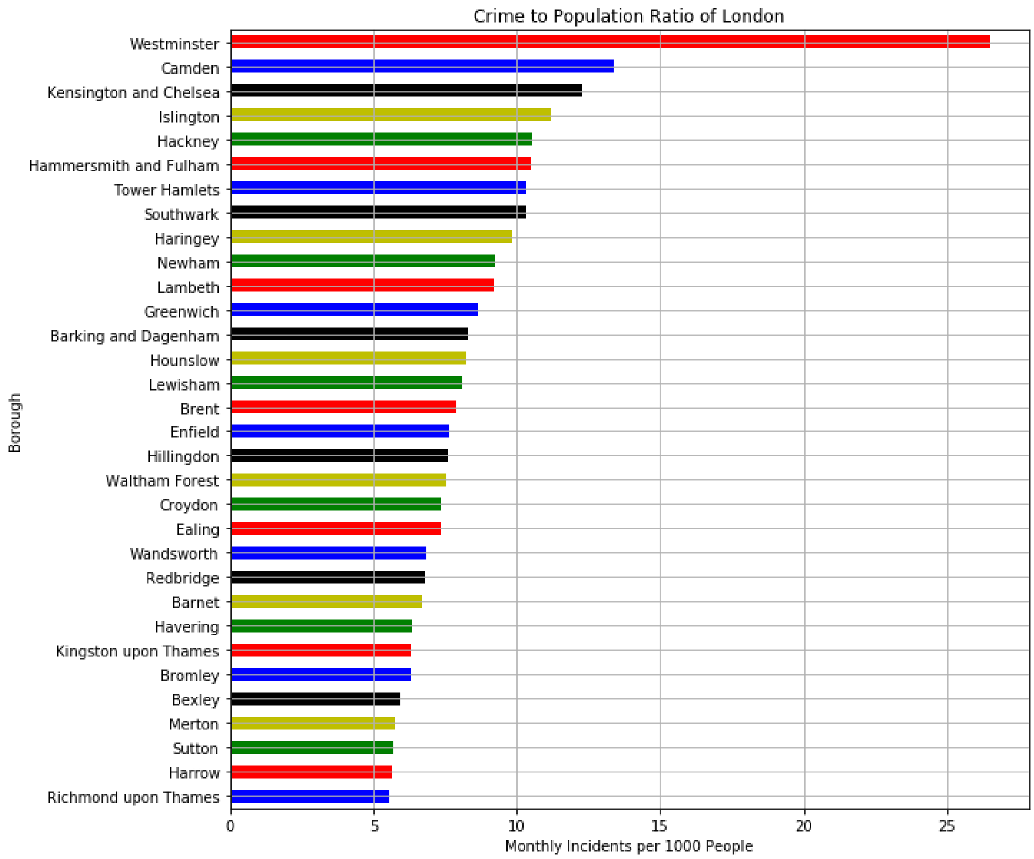
Foursquare API

Foursquare API is used to extract venues from selected neighbourhood

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barking and Dagenham	51.5607	0.1557	Central Park	51.559560	0.161981	Park
1	Barking and Dagenham	51.5607	0.1557	Crowlands Heath Golf Course	51.562457	0.155818	Golf Course
2	Barking and Dagenham	51.5607	0.1557	Robert Clack Leisure Centre	51.560808	0.152704	Martial Arts Dojo
3	Barking and Dagenham	51.5607	0.1557	Morrisons	51.559774	0.148752	Supermarket
4	Barking and Dagenham	51.5607	0.1557	Beacontree Heath Leisure Centre	51.560997	0.148932	Gym / Fitness Center

Methodology

After cleaning and merging part 1 and part 2, the data has been visualised as below. Since population is different for the boroughs, number of recorded crimes for 1000 people has been calculated.

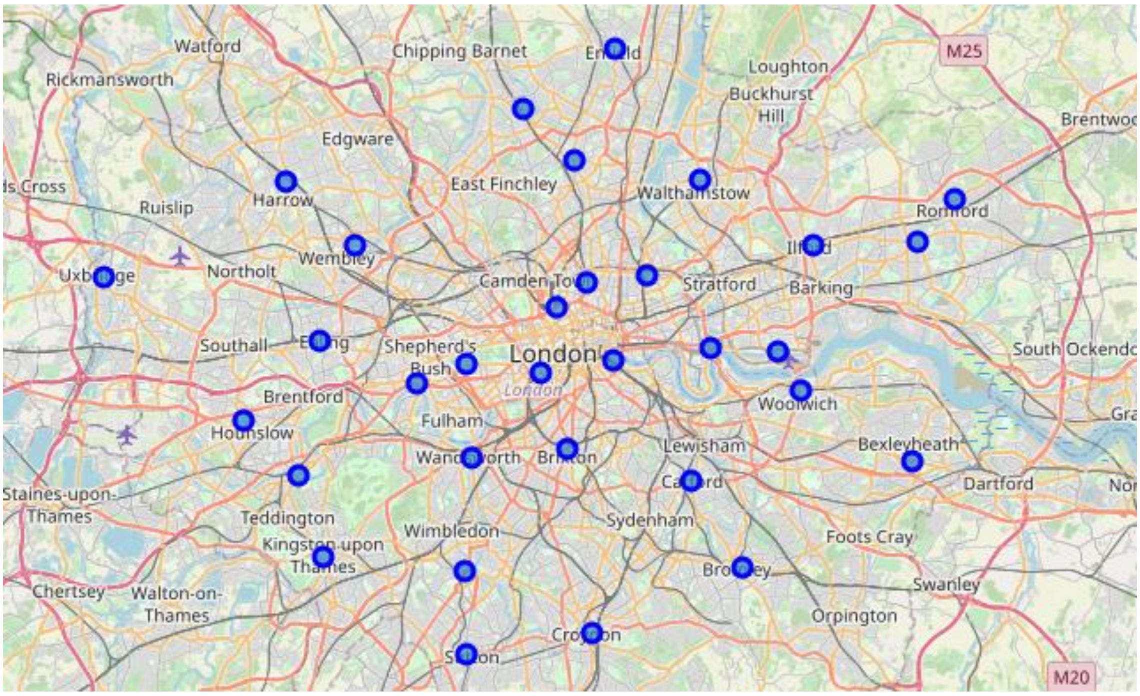


Number of recorded crimes for 1000 people has been calculated as below

```python
#Create a column that shows the number of crimes per 1000 people per month
df_combined['Population'].astype(float)
df_combined['CrimeToPop'] = df_combined['MonthlyAverage'] / df_combined['Population'] * 1000
df_combined.head()
```

	Borough	MonthlyAverage	Population	Latitude	Longitude	CrimeToPop
0	Barking and Dagenham	1613.041667	194352	51.5607	0.1557	8.299589
1	Barnet	2478.500000	369088	51.6252	-0.1517	6.715201
2	Bexley	1409.666667	236687	51.4549	0.1505	5.955826
3	Brent	2508.041667	317264	51.5588	-0.2817	7.905220
4	Bromley	1997.083333	317899	51.4039	0.0198	6.282132

Then, the following map which contains location of each borough has been created



Top 5 venues for each borough has been extracted by using Foursquare API and assigned into a data frame and created a data frame contains most common venues for each borough.

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Barking and Dagenham	Pool	Golf Course	Bus Station	Supermarket
1	Barnet	Café	Bus Stop	Yoga Studio	English Restaurant
2	Bexley	Coffee Shop	Pub	Clothing Store	Fast Food Restaurant
3	Brent	Coffee Shop	Hotel	Clothing Store	Grocery Store
4	Bromley	Clothing Store	Coffee Shop	Burger Joint	Pizza Place

Cluster Analysis

K-Means cluster analysis has been conducted as below to group the boroughs according to the Foursquare data to understand the atmosphere of each borough.

```
# K Means Clustering
from sklearn.cluster import KMeans
import matplotlib.cm as cm
import matplotlib.colors as colors
```

```
kclusters = 5
london_cluster = df_grouped.drop('Borough', 1)
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(london_cluster)
kmeans.labels_
```

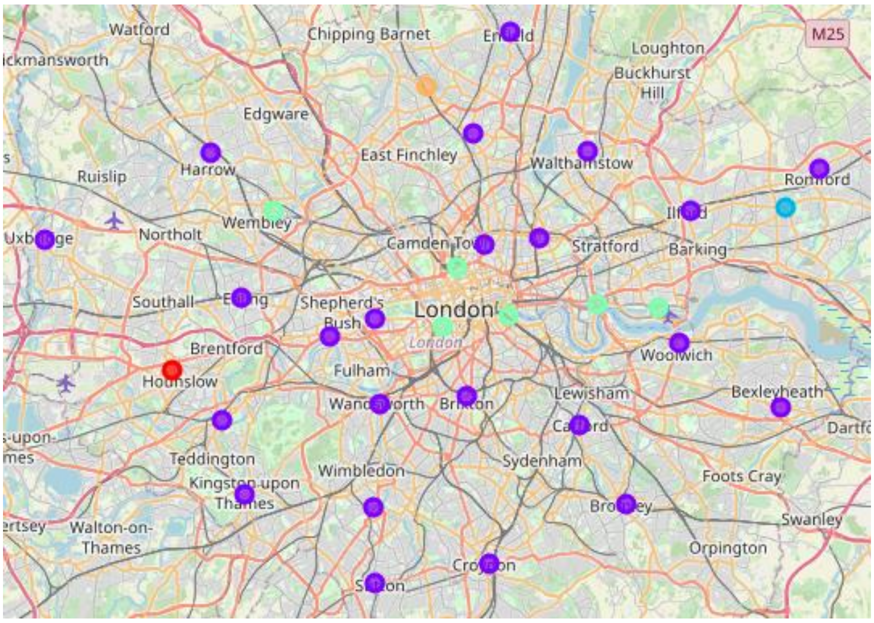
```
array([2, 4, 1, 3, 1, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1,
       1, 3, 1, 1, 3, 1, 3, 1, 1, 3], dtype=int32)
```

```
# add clustering labels
df_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
# merge london_grouped with LONDON coords to add latitude/longitude for each neighborhood
df_london_merged = df_combined
df_london_merged = df_london_merged.join(df_venues_sorted.set_index('Borough'), on='Borough')
df_london_merged.head()
```

The data has been merged with venues data frame as below and created cluster labels for each borough.

	Borough	MonthlyAverage	Population	Latitude	Longitude	CrimeToPop	Cluster Labels
0	Barking and Dagenham	1616.083333	194352	51.5607	0.1557	8.315239	1
1	Barnet	2494.875000	369088	51.6252	-0.1517	6.759567	2
2	Bexley	1412.791667	236687	51.4549	0.1505	5.969029	3
3	Brent	2524.333333	317264	51.5588	-0.2817	7.956570	3
4	Bromley	2009.791667	317899	51.4039	0.0198	6.322108	3

Based on the venue category similarity, 5 different clusters have been created as below map



Clustered venue data has been merged with the crime data as below.

	Borough	CrimeToPop	MonthlyAverage	Population	Latitude	Longitude	Cluster Labels
25	Richmond upon Thames	5.532621	1058.750000	191365	51.4479	-0.3260	1
13	Harrow	5.652526	1375.666667	243372	51.5898	-0.3346	1
27	Sutton	5.705309	1117.750000	195914	51.3618	-0.1945	1
22	Merton	5.764398	1171.458333	203223	51.4014	-0.1958	1
2	Bexley	5.955826	1409.666667	236687	51.4549	0.1505	1

Results

Based on the analysis, best neighbourhoods have been discovered based on safety and atmosphere indexes.

Calculated safety index as below.

```
df_score = df_london_merged[['Borough','CrimeToPop','Cluster Labels']].copy()
df_score['Safety'] = (df_score['CrimeToPop']-df_score['CrimeToPop'].min())/(df_score['CrimeToPop'].max()-df_score['CrimeToPop'].min())
df_score['Safety'] = (df_score['Safety'] - 1) * -1
df_score['Atmosphere'] = 0
df_score.head()
```

	Borough	CrimeToPop	Cluster Labels	Safety	Atmosphere
0	Barking and Dagenham	8.299589	2	0.867976	0
1	Barnet	6.715201	4	0.943574	0
2	Bexley	5.955826	1	0.979807	0
3	Brent	7.905220	3	0.886793	0
4	Bromley	6.282132	1	0.964238	0

Calculated atmosphere index and merged with safety index

	Borough	CrimeToPop	Cluster Labels	Safety	Atmosphere
0	Barking and Dagenham	8.299589	2	0.867976	0.7
1	Barnet	6.715201	4	0.943574	0.9
2	Bexley	5.955826	1	0.979807	1.0
3	Brent	7.905220	3	0.886793	0.8
4	Bromley	6.282132	1	0.964238	1.0
5	Camden	13.366330	3	0.626219	0.8

Calculated total score based on safety and atmosphere index as below

```python
df_score.drop(['CrimeToPop'], inplace=True, axis=1)
df_score.drop(['Cluster Labels'], inplace=True, axis=1)
df_score['Score'] = df_score['Safety'] + df_score['Atmosphere']
df_score.sort_values(by='Score', ascending = False)
```

Based on score top 10 borough can be seen as below

	Borough	MonthlyAverage	Population	Latitude	Longitude	CrimeToPop	Safety	Atmosphere	Score
25	Richmond upon Thames	1058.750000	191365	51.4479	-0.3260	5.532621	1.000000	1.0	2.000000
13	Harrow	1375.666667	243372	51.5898	-0.3346	5.652526	0.994279	1.0	1.994279
27	Sutton	1117.750000	195914	51.3618	-0.1945	5.705309	0.991760	1.0	1.991760
22	Merton	1171.458333	203223	51.4014	-0.1958	5.764398	0.988941	1.0	1.988941
2	Bexley	1409.666667	236687	51.4549	0.1505	5.955826	0.979807	1.0	1.979807
4	Bromley	1997.083333	317899	51.4039	0.0198	6.282132	0.964238	1.0	1.964238
19	Kingston upon Thames	1054.625000	166793	51.4085	-0.3064	6.322957	0.962290	1.0	1.962290
14	Havering	1537.791667	242080	51.5812	0.1837	6.352411	0.960884	1.0	1.960884
24	Redbridge	1966.250000	288272	51.5590	0.0741	6.820815	0.938535	1.0	1.938535
30	Wandsworth	2123.583333	310516	51.4567	-0.1910	6.838885	0.937672	1.0	1.937672

The following map which indicates top 10 boroughs has been created as below

Conclusion

According to the analysis, it has been found that the ten boroughs below are the best places to live based on safety and atmosphere of the neighbourhood.

The following table is sorted based on CrimeToPop, in other words safety index

	Borough	MonthlyAverage	Population	Latitude	Longitude	CrimeToPop	Safety	Atmosphere	Score
25	Richmond upon Thames	1058.750000	191365	51.4479	-0.3260	5.532621	1.000000	1.0	2.000000
13	Harrow	1375.666667	243372	51.5898	-0.3346	5.652526	0.994279	1.0	1.994279
27	Sutton	1117.750000	195914	51.3618	-0.1945	5.705309	0.991760	1.0	1.991760
22	Merton	1171.458333	203223	51.4014	-0.1958	5.764398	0.988941	1.0	1.988941
2	Bexley	1409.666667	236687	51.4549	0.1505	5.955826	0.979807	1.0	1.979807
4	Bromley	1997.083333	317899	51.4039	0.0198	6.282132	0.964238	1.0	1.964238
19	Kingston upon Thames	1054.625000	166793	51.4085	-0.3064	6.322957	0.962290	1.0	1.962290
14	Havering	1537.791667	242080	51.5812	0.1837	6.352411	0.960884	1.0	1.960884
1	Barnet	2478.500000	369088	51.6252	-0.1517	6.715201	0.943574	0.9	1.843574
24	Redbridge	1966.250000	288272	51.5590	0.0741	6.820815	0.938535	1.0	1.938535

The following table is sorted based on Score which is safety index + atmosphere index

	Borough	MonthlyAverage	Population	Latitude	Longitude	CrimeToPop	Safety	Atmosphere	Score
25	Richmond upon Thames	1058.750000	191365	51.4479	-0.3260	5.532621	1.000000	1.0	2.000000
13	Harrow	1375.666667	243372	51.5898	-0.3346	5.652526	0.994279	1.0	1.994279
27	Sutton	1117.750000	195914	51.3618	-0.1945	5.705309	0.991760	1.0	1.991760
22	Merton	1171.458333	203223	51.4014	-0.1958	5.764398	0.988941	1.0	1.988941
2	Bexley	1409.666667	236687	51.4549	0.1505	5.955826	0.979807	1.0	1.979807
4	Bromley	1997.083333	317899	51.4039	0.0198	6.282132	0.964238	1.0	1.964238
19	Kingston upon Thames	1054.625000	166793	51.4085	-0.3064	6.322957	0.962290	1.0	1.962290
14	Havering	1537.791667	242080	51.5812	0.1837	6.352411	0.960884	1.0	1.960884
24	Redbridge	1966.250000	288272	51.5590	0.0741	6.820815	0.938535	1.0	1.938535
30	Wandsworth	2123.583333	310516	51.4567	-0.1910	6.838885	0.937672	1.0	1.937672

The top ten boroughs which is based on Score index all belong to the Lively Area cluster, with many pubs, restaurants, coffee shops and clothing stores. Moreover, the boroughs have high safety index.

References

- “London Recorded Crime: Geographic Breakdown”, London Datastore
- “List of London Boroughs”, Wikipedia
- Foursquare API
- IBM Professional Data Science Specialization Lecture Notes, Coursera