Predicting Socio-Economic Factors impacting the Crime Rates in London Boroughs

Vipin Selvaraj PV

28th March 2019

A. Introduction

A.1. Description & Discussion of the Background

London is a densely populated city with 8.85 million people as of 2018. It is an ever growing city and as per the Office for National Statistics (ONS), London could grow on an average by some 117,000 pa to 9.37 mil in 2021. There are 33 boroughs in London each of which have different characteristics and governance processes. As expected, these have great impact on the socio-economic landscape which poses an unprecedented amount of pressure on Government in order to ingest funds to maintain the quality of life in London. The most concerning issue in every London Borough currently is the growth in the crime rates in recent years.

A.2 Problem

With the confusions over the Brexit looming over the future of UK, and funding cuts affecting the Police, NHS and local government initiatives including the social care, it has become challenging for many boroughs to control the crime rates down.

The factors which could attribute to the crime rates are as follows.

- Population density and degree of urbanization
- Housing
- Poverty
- Social Care
- Family conditions with respect to divorce and family cohesiveness.
- Unemployment
- Effective strength of law enforcement agencies.
- Education
- NHS (National Health Service) - Medical facilities and the financial challenges
- Youth service budget cuts and closures

- Variations in composition of the population, particularly youth concentration.
- Stability of the population with respect to residents' mobility, commuting patterns, and
- Cultural factors and educational, recreational, and religious characteristics.
- Administrative and investigative emphases of law enforcement.
- Citizens' attitudes toward crime.

A.3 Interest

Obviously, the Government would be interested in the accurate prediction of the reasons why the crime rate is increasing, irrespective of all the valuable efforts from their side. The population of the city would also be interested in seeing the crime rates down in London

A.4. Data Description

The datasets I have used for the analysis encompass the following

- Crime Rates in London boroughs

- Socio-Economic Factors : -
  Child Poverty Percentage
  Out of Work Benefit Percentage
  Unemployment Rate
  Family issues (Lone Parents)
  Income Deprived Families
  Reduction in Funding for Youth services
  Child Dropouts from Schools
  Affordable Housing
  The following are the source of data used for analysis.

1 https://data.london.gov.uk
2 https://www.trustforlondon.org.uk/data/
3 https://www.ons.gov.uk
4 https://www.met.police.uk
5 https://www.gmblondon.org.uk
6 https://en.wikipedia.org
7 https://foursquare.com

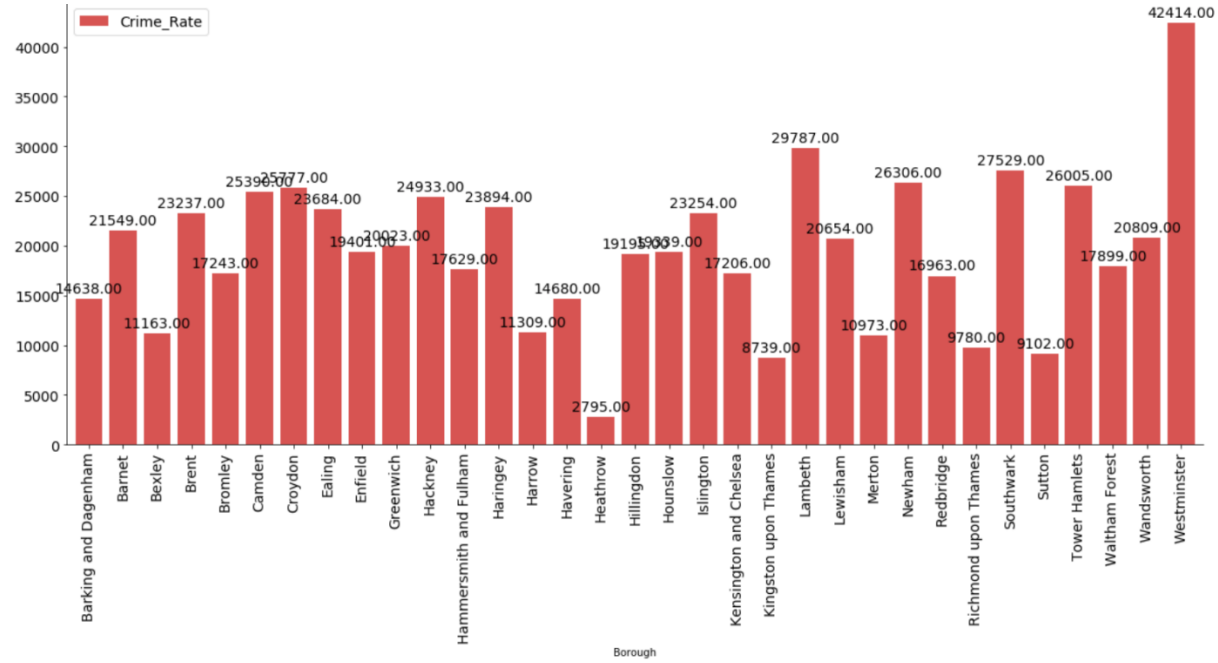
B. Methodology and Analysis

Folium library is used for plotting the heat map of crime rates in London Boroughs K-Means has been used as the cluster analysis Elbow method is used for finding

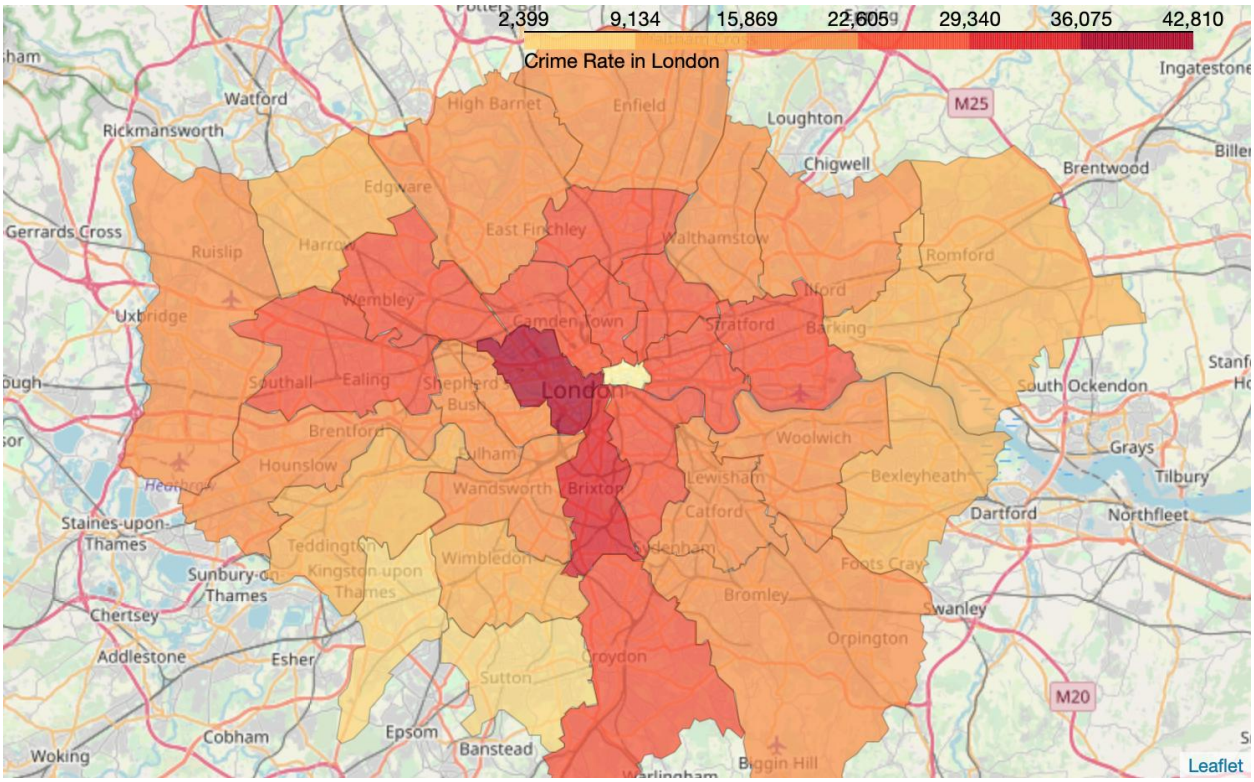
out the optimum number of clusters Pearson Correlation has been used to find out the linear correlation between different factors

Master data for the analysis constitute boroughs, latitude and longitude of boroughs and crime rates associated with each borough. Please find references section for source data details of socio-economic factors.

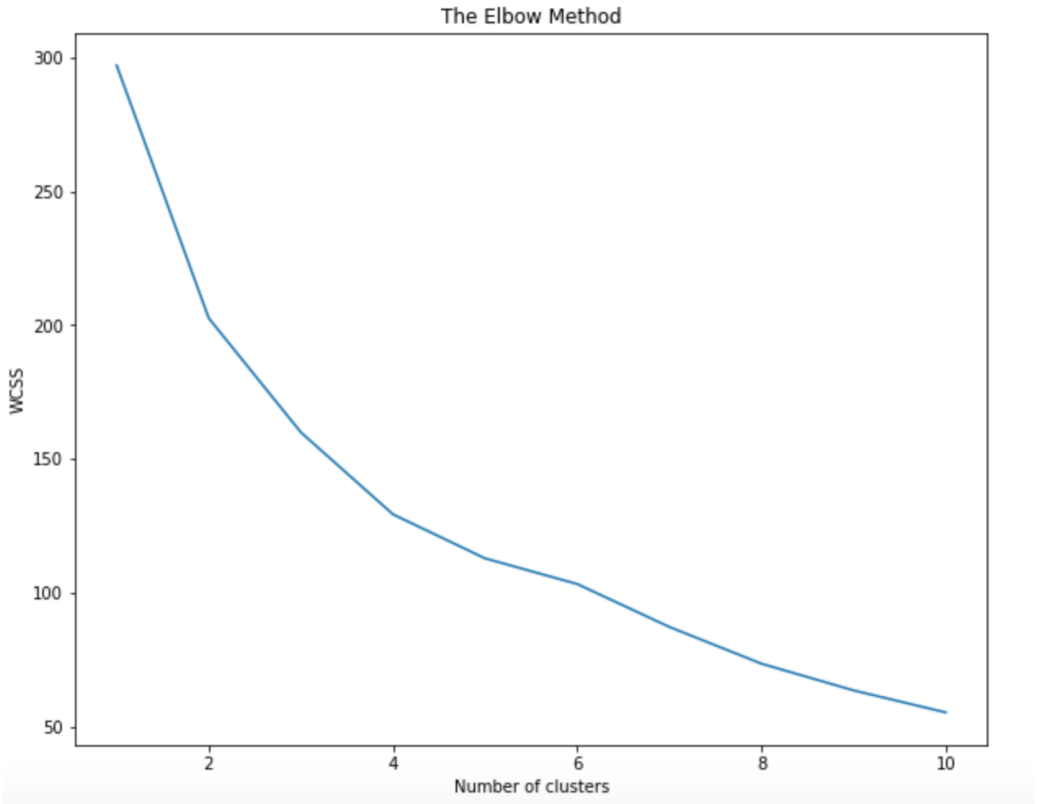
The following is the graph on crime rates across London boroughs

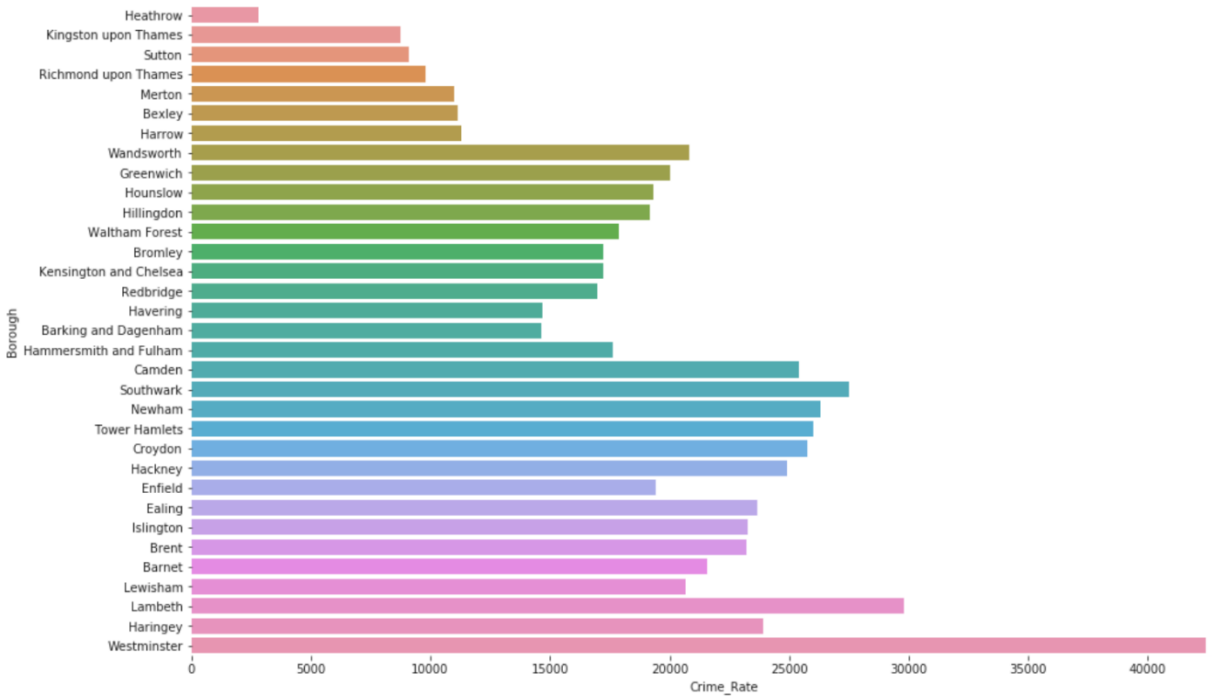


It is evident from the heat map given below that some of the boroughs have exceptionally high crime rates.



Elbow method is used to find out the number of optimum clusters, which is 4



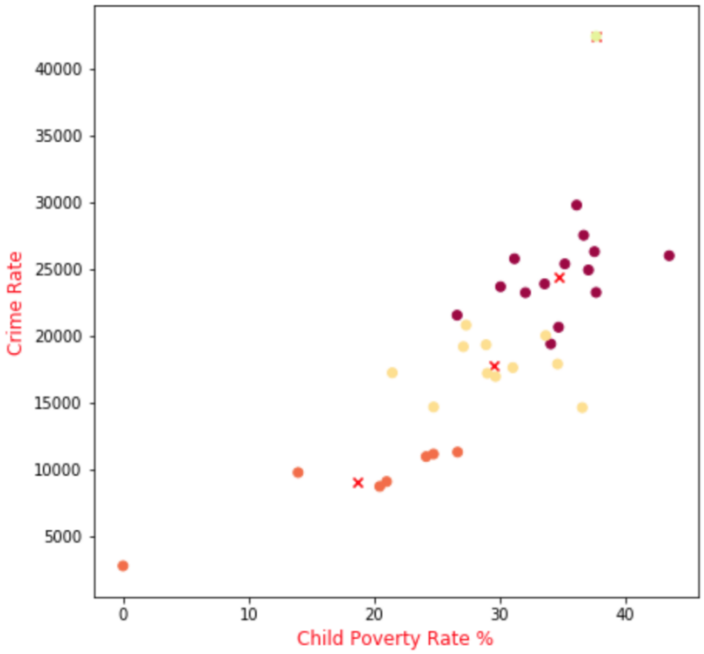


We have 4 optimum clusters. Also it is noticeable that one of the clusters has only one borough, Westminster which has got the highest crime rate.

B.1 K-Means Cluster Analysis

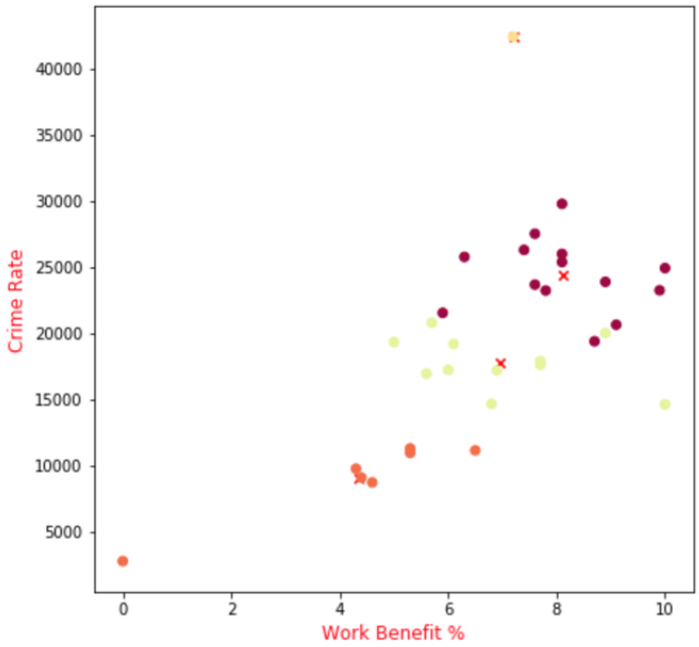
As part of this exercise crime rate is equated against each factor individually to understand the trend of progression and nature of each cluster.

B1.1 Child Poverty and Crime Rate



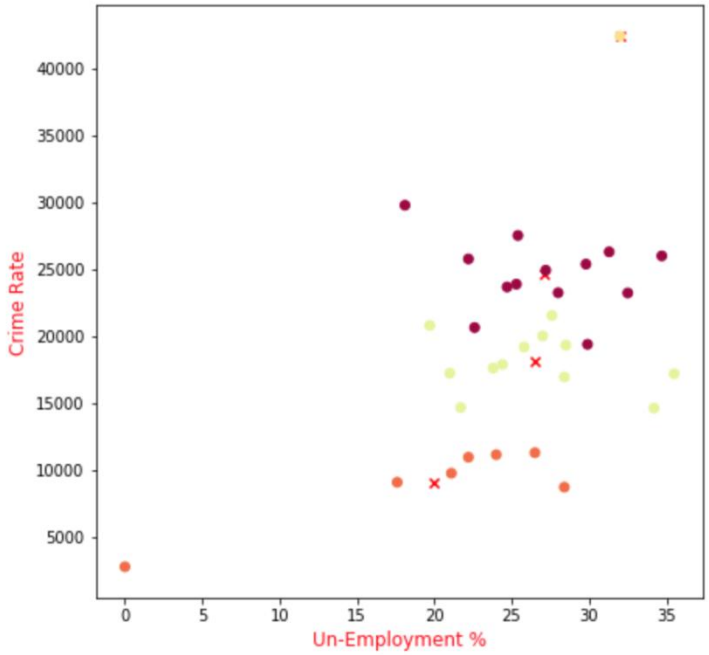
It is evident from the scatter plot that as 'Child Poverty Percentage' increases 'Crime Rate' also goes up sharply. The cluster elements are noticeably concentrated around the centroid.

B1.2 Work Benefit Percentage And Crime Rate



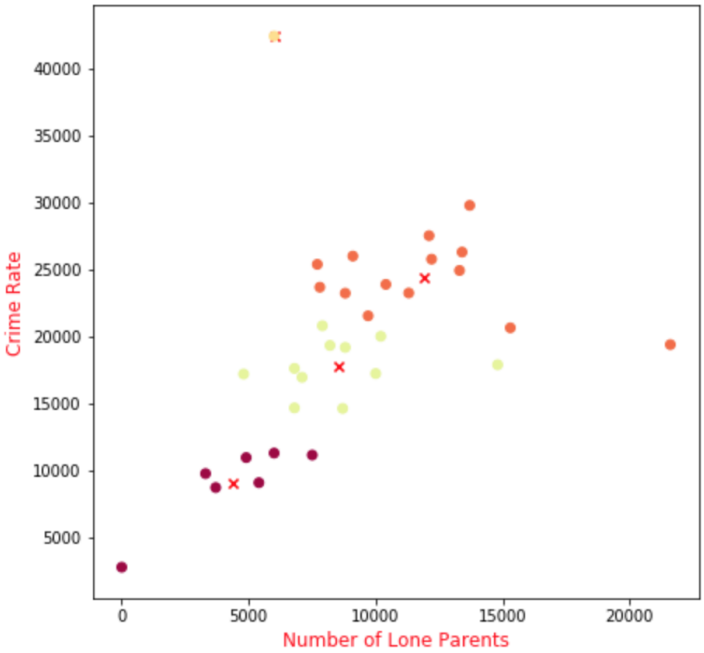
There is a positive correlation between Work Benefit Percentage and Crime Rate. Also there is noticeable flattening of the cluster for the same crime rate.

B1.3 Un-Employment Percentage and Crime Rate



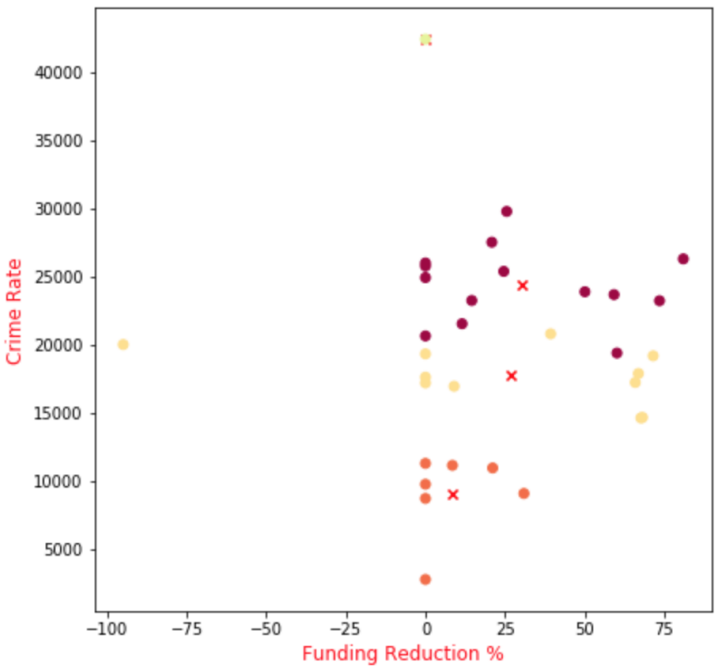
The scatter plot shows a minimal positive increase of Crime Rate with Un-employment percentage

B1.4 Number of Lone Parents and Crime Rate



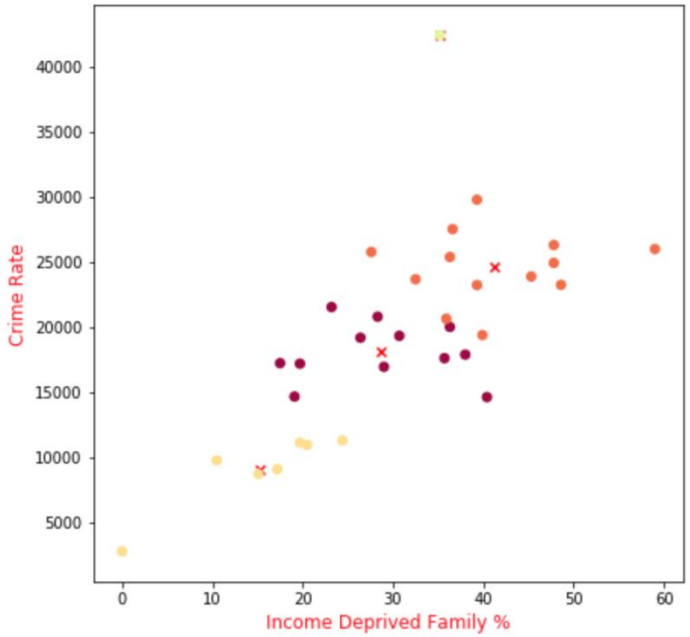
It is evident from the scatter plot that with the increase in the number of lone parents, there is a sharp increase in Crime Rates.

B1.5 Reduction in Funds for Youth Programs and Crime Rate



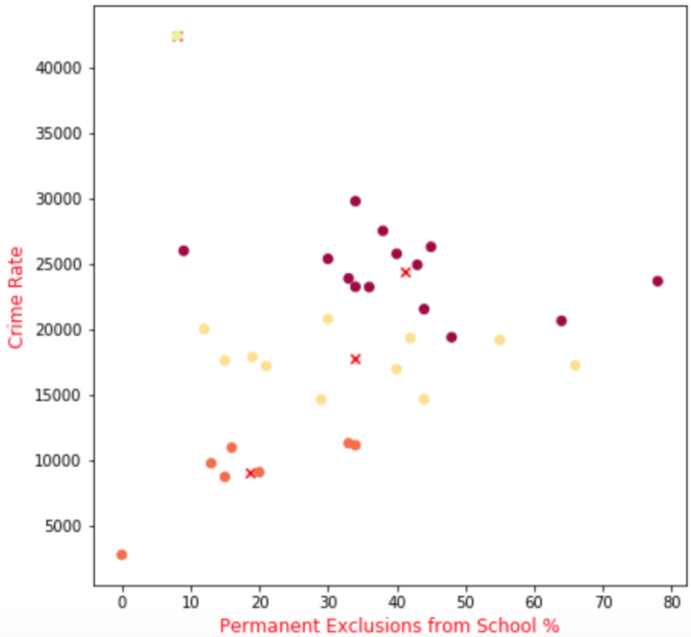
The scatter plot shows that the fund reduction for the Youth Programs has very limited impact on Crime Rates.

B1.6 Income Deprived Family Percentage and Crime Rate



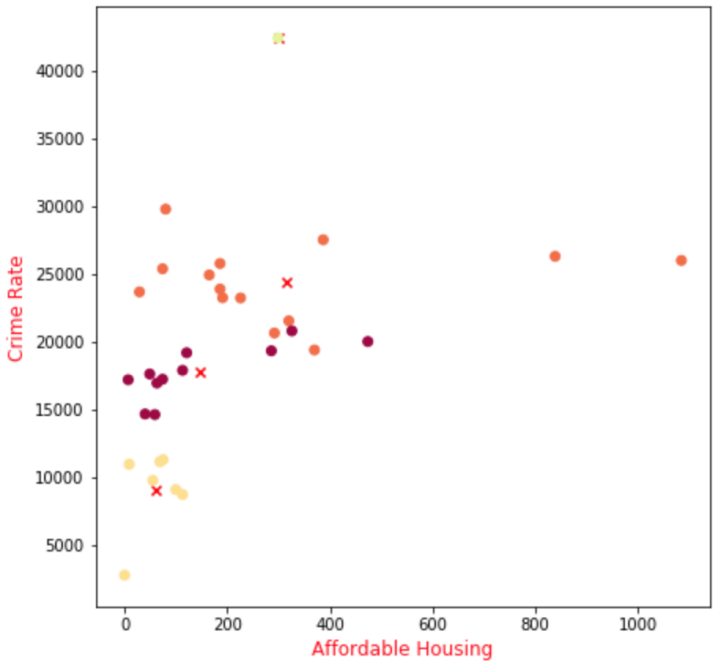
The scatter plot shows that the Percentage of the Income Deprived Families directly impacts the Crime Rates.

B1.7 Permanent Exclusions from School on Crime Rate



The scatter plot shows varied relationship between Percentage exclusions from school and Crime Rate. There is a noticeable amount of flattening of the cluster on the graph.

B1.7 Affordable Housing and Crime Rate



The scatter plot shows varied relationship between Affordable Housing and Crime Rate.

From the K-Means cluster analysis; we could see the positive trend of crime rates with the following factors

- Child Poverty Rate
- Work Benefit Percentage
- Un-Employment Percentage
- Lone Parents
- Income Deprived Families

Let us compare the results with the outcome from Pearson correlation analysis

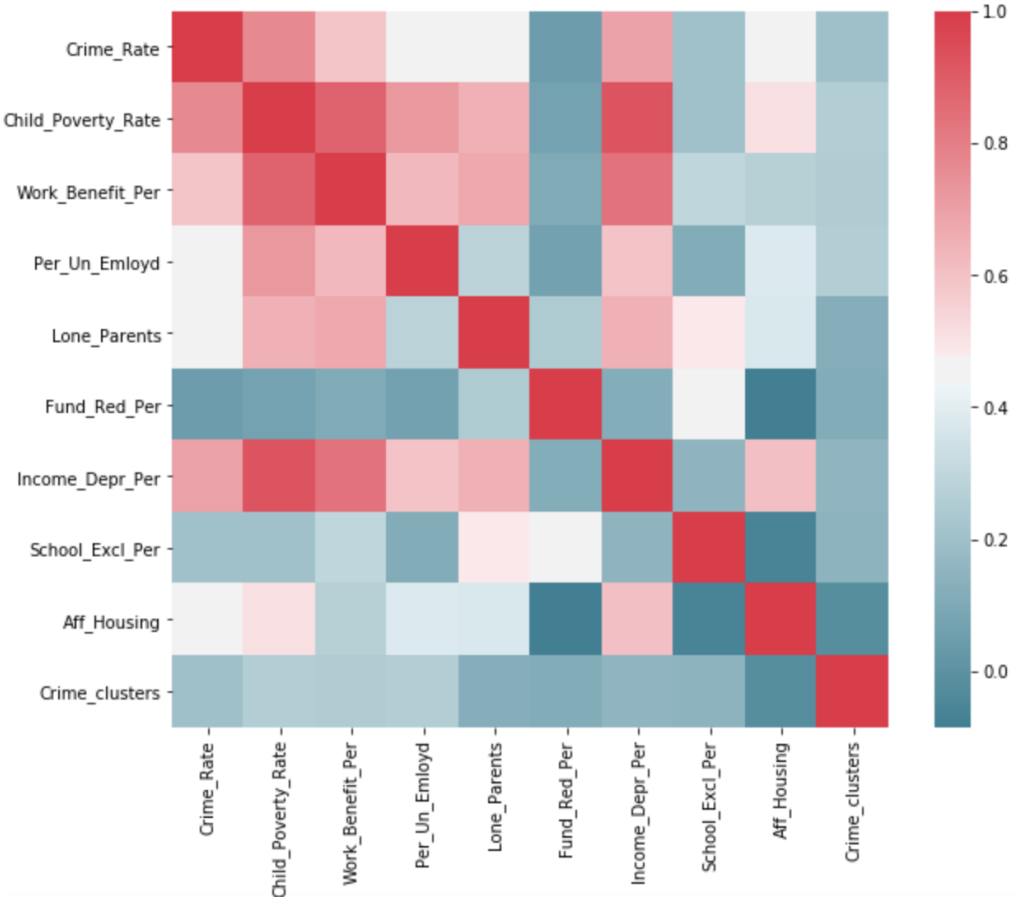
B.2 Pearson Correlation Analysis

The following is the result from Pearson correlation analysis.

	Crime_Rate	Child_Poverty_Rate	Work_Benefit_Per	Per_Un_Emloyd	Lone_Parents	Fund_Red_Per	Income_Depr_Per	School_Excl_Per	Aff_Hous
Crime_Rate	1.000000	0.769093	0.591704	0.457529	0.475792	0.048242	0.691538	0.206605	0.440
Child_Poverty_Rate	0.769093	1.000000	0.879392	0.726126	0.652583	0.071250	0.924661	0.207639	0.511
Work_Benefit_Per	0.591704	0.879392	1.000000	0.625896	0.676641	0.105346	0.837710	0.300892	0.270
Per_Un_Emloyd	0.457529	0.726126	0.625896	1.000000	0.282307	0.065416	0.593898	0.113794	0.383
Lone_Parents	0.475792	0.652583	0.676641	0.282307	1.000000	0.247358	0.649883	0.487268	0.375
Fund_Red_Per	0.048242	0.071250	0.105346	0.065416	0.247358	1.000000	0.114891	0.466493	-0.084
Income_Depr_Per	0.691538	0.924661	0.837710	0.593898	0.649883	0.114891	1.000000	0.151849	0.602
School_Excl_Per	0.206605	0.207639	0.300892	0.113794	0.487268	0.466493	0.151849	1.000000	-0.059
Aff_Housing	0.440820	0.511180	0.270760	0.383762	0.375904	-0.084330	0.602294	-0.059262	1.000
Crime_clusters	0.203115	0.258995	0.251771	0.262698	0.123462	0.112394	0.157023	0.142863	-0.020

It is quite evident that 'Child Poverty Rate' has highest correlation value of 0.769 with crime rate. Income deprived family percentage and work benefit percentage follows child poverty rate. This is in confirmation with what we have found in cluster analysis.

The following is the 'Correlation Heat Map' of the data aforementioned.



C. Results

We have the following final set of data which we have used for analysis.

	Borough	Crime_Rate	Child_Poverty_Rate	Work_Benefit_Per	Per_Un_Emloyd	Lone_Parents	Fund_Red_Per	Income_Depr_Per	School_Excl_Per	Aff_Housing
0	Barking and Dagenham	14638.0	36.57	10.0	34.2	8700.0	67.710611	40.4	29.0	59.0
1	Barnet	21549.0	26.60	5.9	27.6	9700.0	11.503570	23.2	44.0	320.0
2	Bexley	11163.0	24.72	6.5	24.0	7500.0	8.474576	19.7	34.0	69.0
3	Brent	23237.0	32.04	7.8	32.5	8800.0	73.542164	39.3	36.0	226.0
4	Bromley	17243.0	21.44	6.0	21.0	10000.0	65.928440	17.5	66.0	74.0

Number of optimized clusters – 4

The following are the cluster details.

Cluster 0 - Hillingdon, Hounslow, Greenwich… (Third highest crime rate)
Cluster 1 - Westminster (Highest crime rate)
Cluster 2 - Tower of Hamlets, Southwark, Newham … (Second highest crime rate)
Cluster 3 - Sutton, Richmond upon Thames, Harrow…. (Lowest crime rate)

As discussed in the previous section, ‘Child Poverty Rate’, ‘Income Deprived Family Percentage’, ‘Work Benefit Percentage’ top the three main factors which greatly impact the crime rate in London boroughs.

D. Discussion

As discussed before, the population is growing in London and along with the same the crime rate. It is a complex process to analyse all factors affecting the crime rate. Data set for the analysis has been selected with due diligence so as to arrive at accurate results. Also, it was difficult to get data complete data set belonging to all boroughs for some of the parameters considered.

After the data collection, data has gone through validation and cleansing.
As part of the analysis, Folium is used for the display of heat map.

K-means is used for the cluster analysis. I have identified 4 as the optimum number of clusters using the elbow method. Crime rate has been analysed with each factor

to find out the trend of variation. Once the analysis is done, the results are compared with the results from Pearson correlation analysis. They both provided the conclusive results.

E. Conclusion

My analysis reveals that there are mainly 3 groups of socio-economic factors which are affecting the crime rates in London Boroughs

Group 1 - High Impact

1 Child Poverty Rate (Highest impact)
2 Income Deprieved families
3 Work Benefits

Group 2 - Medium Impact

1 Lone Paents
2 Unemployment percentage
3 Un-affordable houses

Group 3 - Low Impact

1 Youth fund reduction
2 School exclusions