

Capstone Project -The Battle of Neighbourhoods

RELOCATING TO LONDON WITH FAMILY – FIND THE BEST AREAS

8 December 2019

Report prepared by: Ajit



INDEX

- Introduction & Business Problem
- Data Description and its Utilisation
- Methodology – Data Analysis
- Data Analysis Results
- Observations and Recommendations
- Conclusion

Introduction & Business Problem

- Capital of UK and Financial Capital of Europe, London is a city of just over 9 million and magnet for many families to move in to start a new career and life. Along with best educational institutes attracting global students, business and employment opportunities attract global talent. Around 30% or 1.1 million of EU nationals living in UK reside in London. Work, Family and Studies combine together to give around 90% as the prime reason for relocation.
- Relocating to London with family is an uphill task which includes finding right place which is safe, commutable to workplace, proximity to schooling, shopping, sports and recreational facilities. Another dimension which plays an important role is cost of accommodation which might be purchase or rent.
- Searching for perfect location which matches all requirements for an individual would be very daunting and time consuming.
- To bridge this gap and cater to relocating client, an Estate agency has come up with this business problem and want to club the additional services together with Property letting/Selling. The aim is to provide client with complete relocation service harnessing power of data science.
- Initial start of the project is to focus on neighbourhood, safety, distance to transport system, schools.

Data Description and Utilisation

Multiple sources of Data are used which have “Boroughs” as a common link.

- Data Source for Borough List with its demographics including Employment, Happiness, Education

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

London Excel Data : <https://data.london.gov.uk/dataset/london-borough-profiles>

- Data Source for London Crime Records – Data Harnessing for Location Safety

Metropliton Police : https://data.london.gov.uk/dataset/recorded_crime_summary

- Four Square Location Data - Location Popularity, neighbourhood, Distances

FourSquare : <https://api.foursquare.com>

The above Data would be systematically filtered, processed and analysed to achieve the desired results.

Methodology – Data Analysis

- Open tables embedded in Webpage are scraped to obtain the data. Columns are dropped, bifurcated and renamed to streamline the data contents. Columns are interpolated to generate additional reference data for Boroughs.
- Wikipedia : https://en.wikipedia.org/wiki/List_of_London_boroughs

wiki.head(4)

	Borough	Population	Area	Latitude	Longitude	Density
0	Barking and Dagenham	194352.0	13.93	51.5607	0.1557	13952.045944
1	Barnet	369088.0	33.49	51.6252	-0.1517	11020.842042
2	Bexley	236687.0	23.38	51.4549	0.1505	10123.481608
3	Brent	317264.0	16.70	51.5588	-0.2817	18997.844311

- Pre-populated official data in spreadsheet form was read from Government web source <https://data.london.gov.uk/dataset/london-borough-profiles> . The Data is cleaned and made compatible to be used in the project.

Lonboro.head(4)

	Borough	Population	Area	Latitude	Longitude	Density	inner_outer_london	gla_household_estimate_2017	average_age_2017	propc
0	Barking and Dagenham	194352.0	13.93	51.5607	0.1557	13952.045944	Outer London	78188.4	32.9	27.2
1	Barnet	369088.0	33.49	51.6252	-0.1517	11020.842042	Outer London	151423	37.3	21.1
2	Bexley	236687.0	23.38	51.4549	0.1505	10123.481608	Outer London	97735.8	39.0	20.6
3	Brent	317264.0	16.70	51.5588	-0.2817	18997.844311	Outer London	121048	35.6	20.9

4 rows x 78 columns

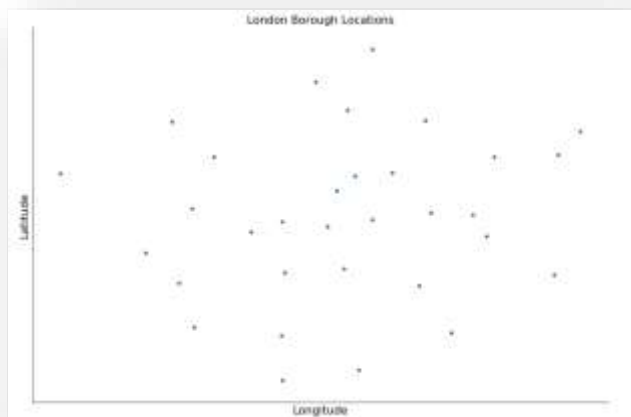
- Explore and Cluster the neighbourhoods in London. Analyse and Report Observations. Visualise to find how the neighbourhoods cluster together. Use of Four Square to Get Venues. Filter the data for Neighbourhood around London Borough.

- Scraped and cleaned Borough data including co-ordinates from Wikipedia in first step is used as basis to acquire data from FourSquare.
- Four Square Location Data - Location Popularity, neighbourhood, Distances

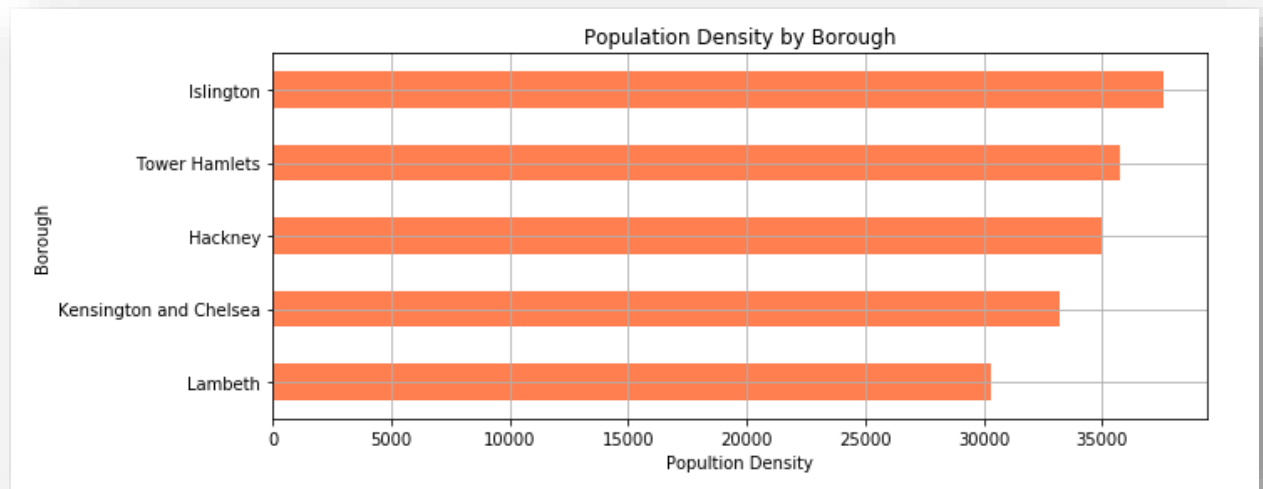
Four-Square : <https://api.foursquare.com>

	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	Beacontree Heath Leisure Centre	51.560997	0.148932	Gym / Fitness Center
4	Barking and Dagenham	51.5607	0.1557	Morrisons Beacontree Heath	51.559774	0.148752	Supermarket

- Cleaned and wrangled data is put to basic queries and interpretation.
- The co-ordinates from different data sources merged together are plotted as a scatter plot. It gives a clear outline of London and its boroughs.



- Sample bar chart is plotted to find the top 5 densely populated boroughs.

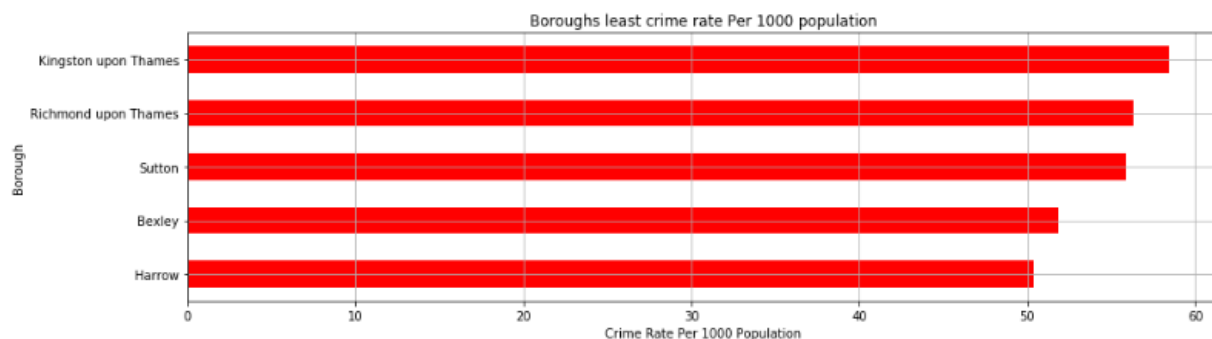


Data Analysis – Results

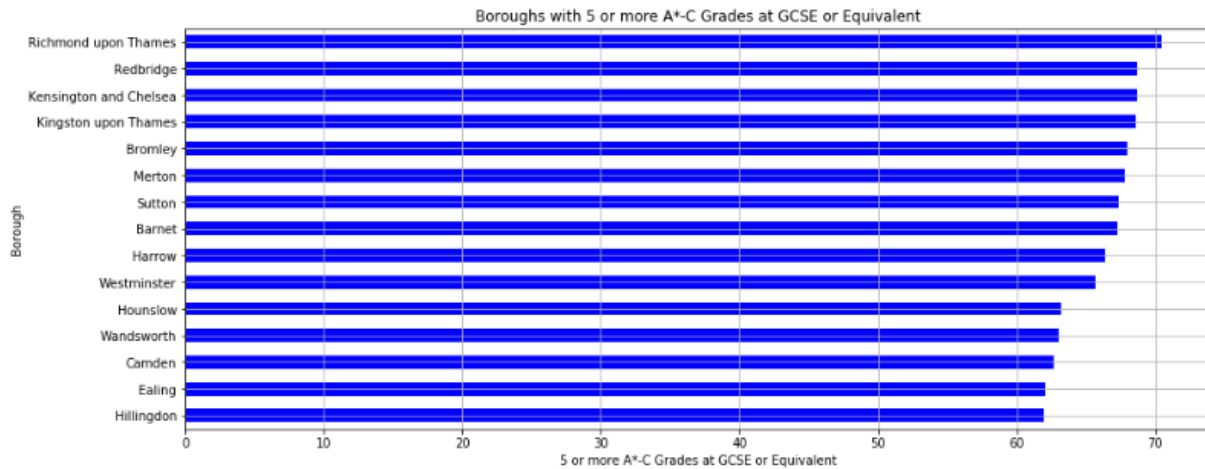
Top 10 frequented venue in each Borough are listed for using Clustering and K-means.

	Borough	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	Barking and Dagenham	Beacontree Heath Leisure Centre	Robert Clark Leisure Centre	Crowlands Heath Golf Course	Dagenham Swimming Pool	Morrisons Beacontree Heath	Central Park	Beacontree Heath Bus Station	Gaucha	Ganleys Irish Bar	Great Northern Hotel
1	Barnet	Beaconstfield Road (BF)	New Barnet Football Academy Ltd	The Atrium	Oakleigh Cafe	Go-Gym	Great Northern Hotel	Granger & Co.	Grange Wellington Hotel	Gourmet Burger Kitchen	Goodlooking Optics
2	Bexley	Marks & Spencer	Sainsbury's	Argos	Prince Albert	Lidl	Boots	Wilko	Nando's	New Look	Bexleyheath Marriott Hotel
3	Brent	Starbucks	Superdry	Arena Square	TGI Fridays	Premier Inn London Wembley Park	Premier Inn London Wembley Stadium	Las Iguanas	Moulin Cafe	Caffè Nero	Cineworld
4	Bromley	Caffè Nero	Costa Coffee	wagamama	PureGym	Paperchase	Patisserie Valerie	Pavilion Leisure Centre	PizzaExpress	Pret A Manger	Primark

- Now the data is ready to provide all combinations of queries. As the project is intended to provide services for families wanting to relocate to London few questions form the basis.
 - Which Boroughs are safest in terms of crime rate?
 - Here we query the data for Boroughs with least crime rate per 1000 people in year?



- Which Boroughs have students achieving higher grades at their GCSE and equivalents.
 - Here we look at top 15 Boroughs which show highest number of students with 5 or more A*-C grades



- Which Borough is with highest un-employment rate?
 - We query the data to get the Borough with highest un-employment rate in terms of %.

%	Borough	unemployment_rate_2015
11.0	Barking and Dagenham	11.0

Observations & Recommendations

Observations

- The top 5 safest boroughs also are part of the list where students have highest percentage of 5plus A*-C grades at GCSE levels.
- These would be the five areas to looked at places to re-locate.
- Filtered locations open up options to query the data further to provide answer to further queries to pin point prominent boroughs of interest.

Recommendations

- The area of Harrow stands out to be an optimum balance satisfying the following requirements
 - Safe location
 - Excellent Educational infrastructure giving high percentage of 5 plus A*-C Students
 - Lower un-employment rates
 - Plenty frequented Recreational and Sports facilities
 - Access to transportation systems. Airport, Trains & Tubes(underground)
 - Moderate housing prices and borough with good per

	Borough	Latitude	Longitude	median_house_price_2015	Employment	Life	Transport	School	Safety	Weight
25	Richmond upon Thames	51.4479	-0.3260	575000	0.62	0.754	0.305746	0.705	0.436522	2.821268
13	Harrow	51.5898	-0.3346	396150	0.54	0.733	0.28702	0.664	0.496039	2.720059
22	Merton	51.4014	-0.1958	415000	0.54	0.728	0.340559	0.678	0.402528	2.689087
19	Kingston upon Thames	51.4085	-0.3064	410000	0.55	0.734	0.292825	0.686	0.415443	2.678268
27	Sutton	51.3618	-0.1945	320000	0.45	0.746	0.285409	0.674	0.441439	2.596847

Conclusion

Using the Data Analysis we have successfully been able to filter down to Top 5 locations where a family can re-locate based on the criteria. Depending on the budget, the housing prices give an indication about the borough which can be preferred.

The current data provides options to search from more than 70 different Demographic criteria.

The interfaces can be customized to be driven through a drop-down selection criterion providing easy interface to end user.

