

An Inside into London through its Boroughs

ABHIMANU SHARMA

AUGUST, 2020

1. Introduction:

- 1.1. Background: In this era of Globalization, borders are limited to maps on the papers. With the increasing movement of peoples across the borders, it becomes essential to get some knowledge about the place where one is planning to move. Different people have different situations and reasons for migrating from one country to another, and thus they have different needs. This project is an attempt to cater to the spectrum of such different needs for the people moving to London, UK. In this project, I try to describe the 32 boroughs of London based on the data of various parameters provided by the UK government. Foursquare API is used to get the information on Venues in the neighborhoods of different boroughs.
- 1.2. Problem: Expats coming to London to settle down should have a broad perspective of the city's demographics, economics, and financial standards of various parts of the city. Without a proper understanding of the place, they may find it challenging to find a place suitable for their settlement with the neighborhood that meets their requirements.
- 1.3. Interest: People of different age groups, gender, occupation status, students, or anyone who is planning to settle down for a considerable amount of time in London will find this piece of information useful in making up their minds and getting familiar with features of various boroughs of London.

2. Data:

- 2.1. Data Sources: For the list of areas in London, the Wikipedia page for London areas, https://en.wikipedia.org/wiki/List_of_areas_of_London was scrapped using BeautifulSoup library of Python. Data regarding the profiles of London Boroughs was obtained from the link <https://old.datahub.io/dataset/london-borough-profiles>, it is the official data provided by the government authorities of the United Kingdom.

2.2. Data Cleaning and Data Preparation: After getting the data from the Wikipedia page, it was then converted to a tabular format into a Pandas Dataframe, which looked like:

	Neighborhood	London borough	Post town	Postcode district
0	Abbey Wood	Bexley, Greenwich	LONDON	SE2
1	Acton	Ealing, Hammersmith and Fulham	LONDON	W3, W4
2	Addington	Croydon	CROYDON	CR0
3	Addiscombe	Croydon	CROYDON	CR0
4	Albany Park	Bexley	BEXLEY, SID CUP	DA5, DA14
5	Aldborough Hatch	Redbridge	ILFORD	IG2
6	Aldgate	City	LONDON	EC3

Before using the above data with the FOURSQUARE API, the Neighborhoods must be linked with their respective Latitude and Longitude. For this purpose, NOMINATIM was used to get the google coordinates for each Neighborhood location, and they were populated to the above Dataframe to give the following result:

	Neighborhood	London borough	Post town	Postcode district	Neighborhood Latitude	Neighborhood Longitude
0	Abbey Wood	Bexley, Greenwich	LONDON	SE2	51.4876	0.11405
1	Acton	Ealing, Hammersmith and Fulham	LONDON	W3, W4	51.5081	-0.273261
2	Addington	Croydon	CROYDON	CR0	51.3586	-0.0316347
3	Addiscombe	Croydon	CROYDON	CR0	51.3797	-0.0742821
4	Albany Park	Bexley	BEXLEY, SID CUP	DA5, DA14	51.4354	0.125965
5	Aldborough Hatch	Redbridge	ILFORD	IG2	51.5855	0.0988

The second set of Data regarding the London Borough Profile was directly downloaded from the website (link provided above). The raw Data looked initially like this:

	Code	Area_name	Inner/_Outer_London	GLA_Population_Estimate_2017	GLA_Household_Estimate_2017	Inland_Area_(Hectares)	Population_density_(per_hectare)_2017	Average_Age_2017
0	E09000001	City of London	Inner London	8800	5326	290	30.3	43
1	E09000002	Barking and Dagenham	Outer London	209000	78188	3,611	57.9	32
2	E09000003	Barnet	Outer London	389600	151423	8,675	44.9	37
3	E09000004	Bexley	Outer London	244300	97736	6,058	40.3	39
4	E09000005	Brent	Outer London	332100	121048	4,323	76.8	35

In this Data, missing values were replaced by NaN values, data-type of columns were converted from object-type to float-type, and finally, Latitude and Longitude columns were created and populated using the GEOLOCATOR-NOMINATIM. Then only columns required for analysis were kept, and the rest were dropped before plotting each feature as Heat-Map.

3. Methodology:

After getting the Data on the Neighborhood venues from FOURSQUARE API, it was grouped by venue categories. There were a total of 410 venue categories found. These 410 venue categories were further divided into 14 different Venue Types, and then these venue-types were again repopulated into the Dataframe:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Type
0	Abbey Wood	51.487621	0.114050	Co-op Food	51.487650	0.113490	Grocery Store	Grocery_daily
1	Abbey Wood	51.487621	0.114050	Bostal Gardens	51.486670	0.110462	Playground	Sports_fitness
2	Abbey Wood	51.487621	0.114050	Abbey Wood Caravan Club	51.485502	0.120014	Campground	Sports_fitness
3	Acton	51.508140	-0.273261	London Star Hotel	51.509624	-0.272456	Hotel	Hotels_boarding
4	Acton	51.508140	-0.273261	The Aeronaut	51.508376	-0.275216	Pub	Pub_bars

The 14 Venue Types and the number of venues in each category is as follows:

	Venue Type	Count
0	Cafe_fastfood	3016
1	Restaurants	2418
2	Pubs_bars	1453
3	Shopping_retail	1154
4	Grocery_daily	882
5	Entertainment_recreation	702
6	Sports_fitness	572
7	Hotels_boarding	436
8	Transportation	432
9	Tourists_landmarks	315
10	Pharmacy	145
11	Personal_care	45
12	Business_office	25
13	Education_school	5

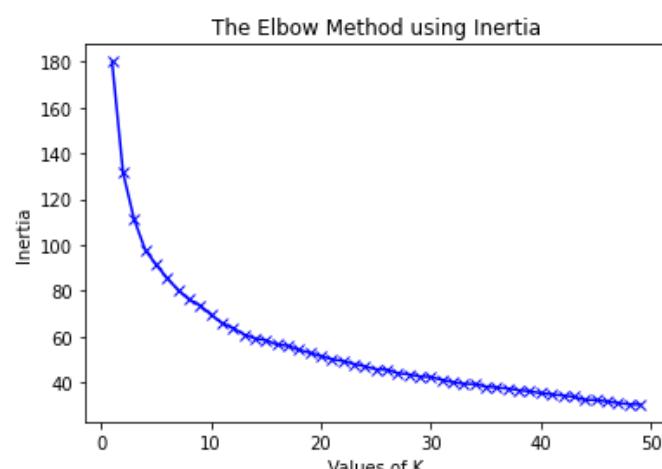
As London is a metropolitan state, the number of restaurants, pubs, bars, and cafes will overpower the other venue types. Thus, weights were assigned to each venue-type to neutralize this bias, which can have a drastically adverse impact on clustering of neighborhoods based on these venue types:

	Venue Type	Count	Weight
0	Business_office	25	0.464000
1	Cafe_fastfood	3016	0.003846
2	Education_school	5	2.320000
3	Entertainment_recreation	702	0.016524
4	Grocery_daily	882	0.013152
5	Hotels_boarding	436	0.026606
6	Personal_care	45	0.257778
7	Pharmacy	145	0.080000
8	Pubs_bars	1453	0.007983
9	Restaurants	2418	0.004797
10	Shopping_retail	1154	0.010052
11	Sports_fitness	572	0.020280
12	Tourists_landmarks	315	0.036825
13	Transportation	432	0.026852

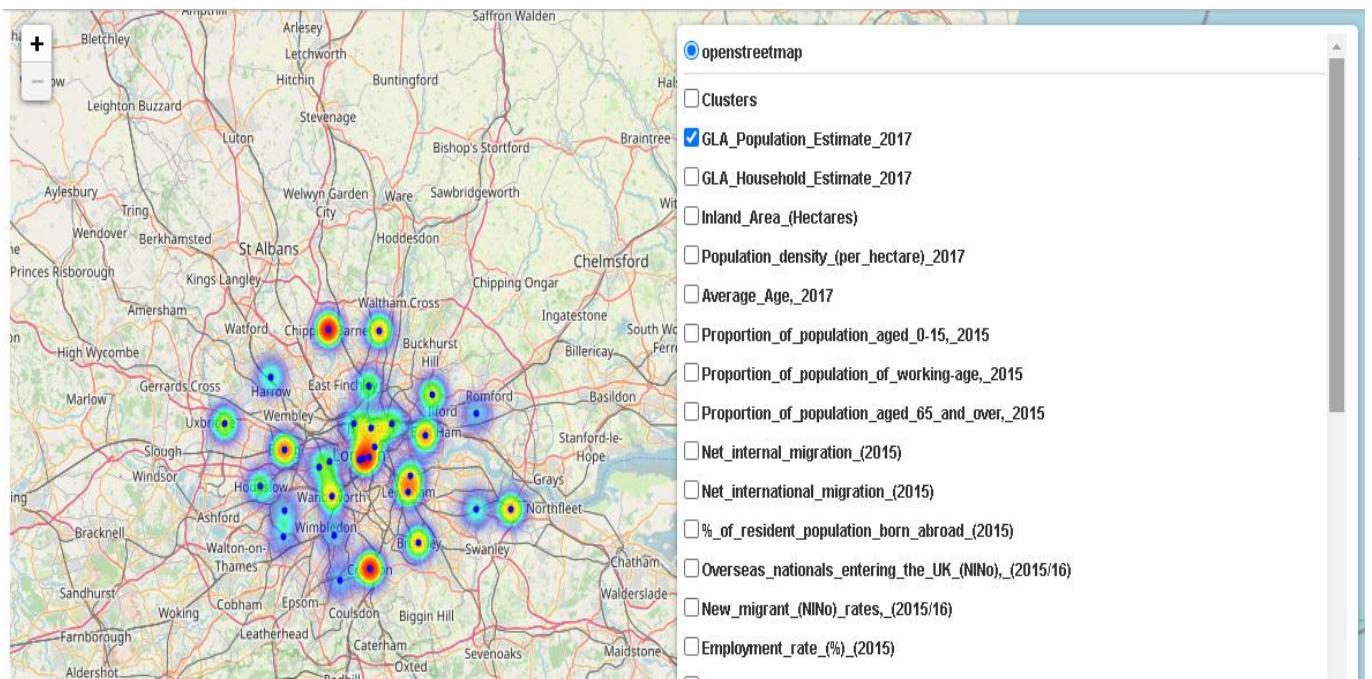
Furthermore, the neighborhoods were one-hot encoded as per the venue-types, and the sum was multiplied by the weights calculated above, thereafter *features were standardized* using Min-Max scaler from Scikit-learn library:

	Neighborhood	Business_office	Cafe_fastfood	Education_school	Entertainment_recreation	Grocery_daily	Hotels_boarding	Personal_care	P
0	Abbey Wood	0.0	0.000	0.0	0.000000	0.125	0.000000	0.0	
1	Acton	0.0	0.100	0.0	0.000000	0.125	0.045455	0.0	
2	Addington	0.0	0.000	0.0	0.066667	0.000	0.000000	0.0	
3	Addiscombe	0.0	0.075	0.0	0.133333	0.250	0.000000	0.0	
4	Albany Park	0.0	0.000	0.0	0.000000	0.125	0.000000	0.0	

The above data prepared for clustering is then used for *K-Means clustering*. *Elbow method* was used to select the appropriate number of clusters (which came out to be TEN):



On the other hand, with the second set of data, Folium was used to create different FeatureGroups for different features for each borough. These FeatureGroups were then finally added '*as_child*' to the base Map of the London State. LayerControl Feature of the Folium library was also added to the base map for flexible and easy interpretation of the borough profiles:



4. RESULT:

After clustering the neighborhood-venues data, ten clusters were formed:

No. of Neighborhoods in Cluster 1:	31
No. of Neighborhoods in Cluster 2:	254
No. of Neighborhoods in Cluster 3:	24
No. of Neighborhoods in Cluster 4:	5
No. of Neighborhoods in Cluster 5:	32
No. of Neighborhoods in Cluster 6:	39
No. of Neighborhoods in Cluster 7:	35
No. of Neighborhoods in Cluster 8:	5
No. of Neighborhoods in Cluster 9:	19
No. of Neighborhoods in Cluster 10:	66

These clusters were then analyzed based on the number of venue types in each cluster. Thereafter, cluster profiling was done, and each cluster was named based on what distinct feature it represents:

Venue Type	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	SUM
Business_office	0	0	0	5	0	0	0	1	19	0	25
Cafe_fastfood	115	375	420	108	302	302	774	89	91	440	3016
Education_school	0	0	0	0	0	0	0	5	0	0	5
Entertainment_recreation	30	144	67	18	87	40	208	16	18	74	702
Grocery_daily	47	177	117	17	48	107	67	9	30	263	882
Hotels_boarding	18	60	53	37	41	24	129	18	16	40	436
Personal_care	4	8	4	1	6	2	12	3	0	5	45
Pharmacy	4	5	44	5	3	59	6	3	4	12	145
Pubs_bars	50	250	170	64	142	123	335	41	55	223	1453
Restaurants	86	320	349	102	240	250	579	70	75	347	2418
Shopping_retail	34	122	254	39	79	133	321	41	40	91	1154
Sports_fitness	30	121	83	13	130	28	66	12	23	66	572
Tourists_landmarks	22	41	15	11	47	8	132	9	15	15	315
Transportation	137	113	50	5	15	26	11	5	21	49	432
TOTAL	577	1736	1626	425	1140	1102	2640	322	407	1625	11600

Percentage Distribution of Venue-Types across Clusters (Row-wise):

Venue Type	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Business_office	0.000000	0.000000	0.000000	20.000000	0.000000	0.000000	0.000000	4.000000	76.000000	0.000000
Cafe_fastfood	3.812997	12.433687	13.925729	3.580902	10.013263	10.013263	25.663130	2.950928	3.017241	14.588859
Education_school	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	100.000000	0.000000	0.000000
Entertainment_recreation	4.273504	20.512821	9.544160	2.564103	12.393162	5.698006	29.629630	2.279202	2.564103	10.541311
Grocery_daily	5.328798	20.068027	13.265306	1.927438	5.442177	12.131519	7.596372	1.020408	3.401361	29.818594
Hotels_boarding	4.128440	13.761468	12.155963	8.486239	9.403670	5.504587	29.587156	4.128440	3.669725	9.174312
Personal_care	8.888889	17.777778	8.888889	2.222222	13.333333	4.444444	26.666667	6.666667	0.000000	11.111111
Pharmacy	2.758621	3.448276	30.344828	3.448276	2.068966	40.689655	4.137931	2.068966	2.758621	8.275862
Pubs_bars	3.441156	17.205781	11.699931	4.404680	9.772884	8.465244	23.055747	2.821748	3.785272	15.347557
Restaurants	3.556658	13.234078	14.433416	4.218362	9.925558	10.339123	23.945409	2.894955	3.101737	14.350703
Shopping_retail	2.946274	10.571924	22.010399	3.379549	6.845754	11.525130	27.816291	3.552860	3.466205	7.885615
Sports_fitness	5.244755	21.153846	14.510490	2.272727	22.727273	4.895105	11.538462	2.097902	4.020979	11.538462
Tourists_landmarks	6.984127	13.015873	4.761905	3.492063	14.920635	2.539683	41.904762	2.857143	4.761905	4.761905
Transportation	31.712963	26.157407	11.574074	1.157407	3.472222	6.018519	2.546296	1.157407	4.861111	11.342593

Percentage Distribution of Venue-Types within each Cluster (Column-wise):

Venue Type	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Business_office	0.000000	0.000000	0.000000	1.176471	0.000000	0.000000	0.000000	0.310559	4.668305	0.000000
Cafe_fastfood	19.930676	21.601382	25.830258	25.411765	26.491228	27.404719	29.318182	27.639752	22.358722	27.076923
Education_school	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.552795	0.000000	0.000000
Entertainment_recreation	5.199307	8.294931	4.120541	4.235294	7.631579	3.629764	7.878788	4.968944	4.422604	4.553846
Grocery_daily	8.145581	10.195853	7.195572	4.000000	4.210526	9.709619	2.537879	2.795031	7.371007	16.184615
Hotels_boarding	3.119584	3.456221	3.259533	8.705882	3.596491	2.177858	4.886364	5.590062	3.931204	2.461538
Personal_care	0.693241	0.460829	0.246002	0.235294	0.526316	0.181488	0.454545	0.931677	0.000000	0.307692
Pharmacy	0.693241	0.288018	2.706027	1.176471	0.263158	5.353902	0.227273	0.931677	0.982801	0.738462
Pubs_bars	8.665511	14.400922	10.455105	15.058824	12.456140	11.161525	12.689394	12.732919	13.513514	13.723077
Restaurants	14.904679	18.433180	21.463715	24.000000	21.052632	22.686025	21.931818	21.739130	18.427518	21.353846
Shopping_retail	5.892548	7.027650	15.621156	9.176471	6.929825	12.068966	12.159091	12.732919	9.828010	5.600000
Sports_fitness	5.199307	6.970046	5.104551	3.058824	11.403509	2.540835	2.500000	3.726708	5.651106	4.061538
Tourists_landmarks	3.812825	2.361751	0.922509	2.588235	4.122807	0.725953	5.000000	2.795031	3.685504	0.923077
Transportation	23.743501	6.509217	3.075031	1.176471	1.315789	2.359347	0.416667	1.552795	5.159705	3.015385

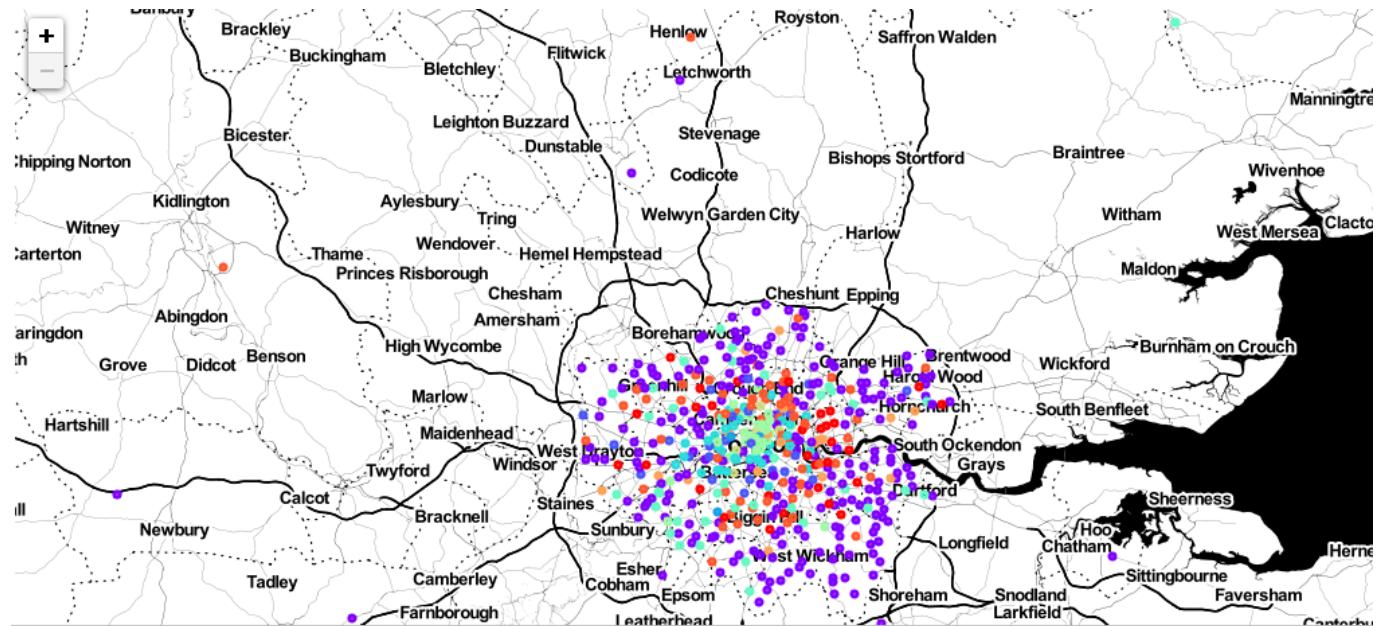
Based on the analysis from the above visualization of the distribution of Venue-Types across and within clusters, 9 out of 10 clusters were profiled. Following profiling has been deduced (**high prevalence of Restaurants, Pubs_bars, and Cafe_fastfood was neglected while profiling the clusters**):

- 1) CLUSTER-1: Transportation Cluster.
- 2) CLUSTER-2: Common Neighborhood Cluster. It contains Grocery, Shopping Retail, etc. in moderation.
- 3) CLUSTER-3: Shopping Retail Cluster.
- 4) CLUSTER-4: Ambiguous, thus, cannot be profiled.
- 5) CLUSTER-5: Sports & Fitness Cluster.
- 6) CLUSTER-6: Pharmacy Cluster.
- 7) CLUSTER-7: Master Cluster, as it contains almost everything in a significant proportion.
- 8) CLUSTER-8: Education Cluster.
- 9) CLUSTER-9: Business & Office Cluster.
- 10) CLUSTER-10: Grocery Daily Cluster along with Cafe, Fastfood, and Restaurants.

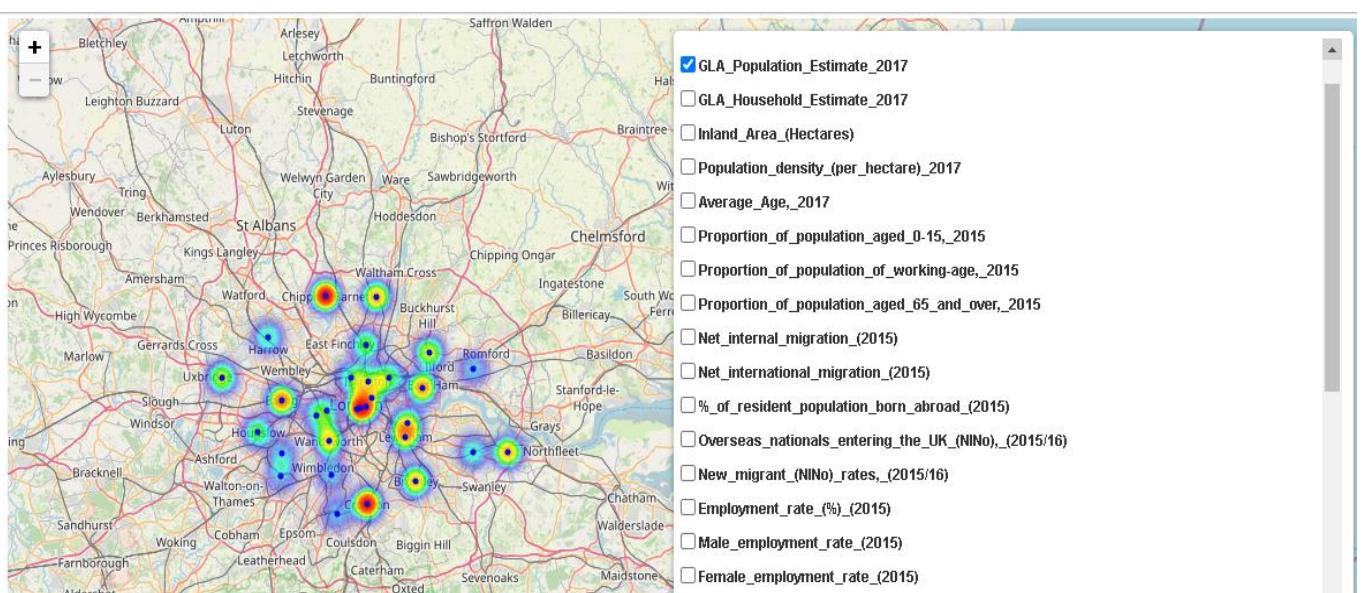
Most neighborhood falls in Cluster-2, and as it the Common neighborhood cluster, these neighborhoods are mostly outside central London.

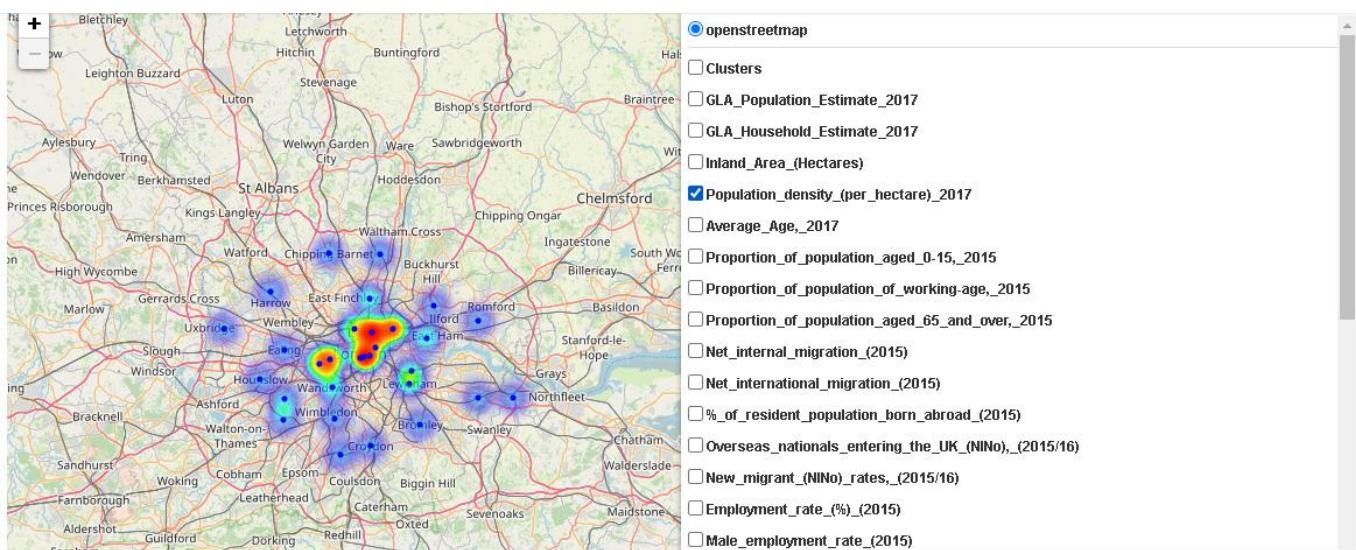
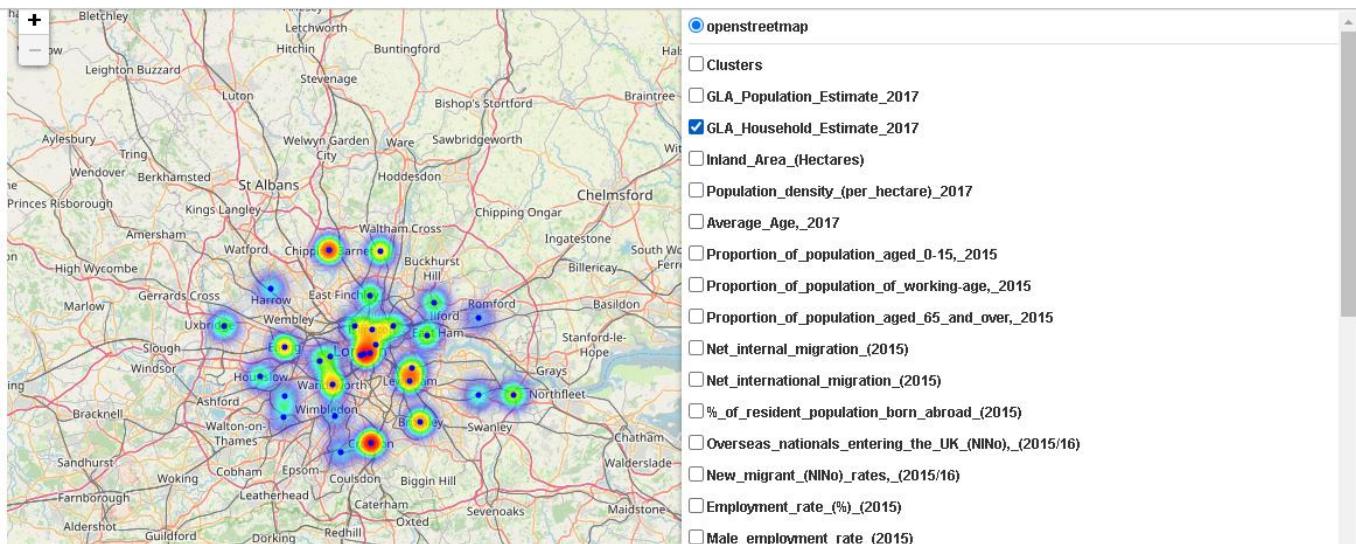
5. Discussion:

The clustering graph, as shown below, along with the Heat-Maps of various features, allows us to deduce some insights regarding various London Boroughs.

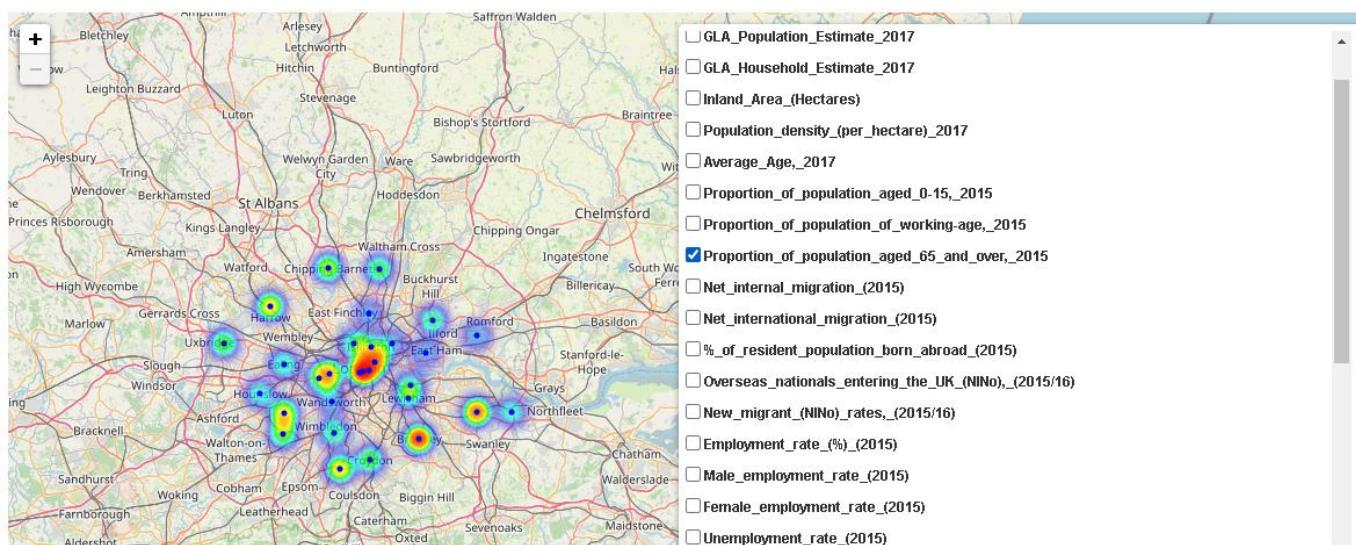


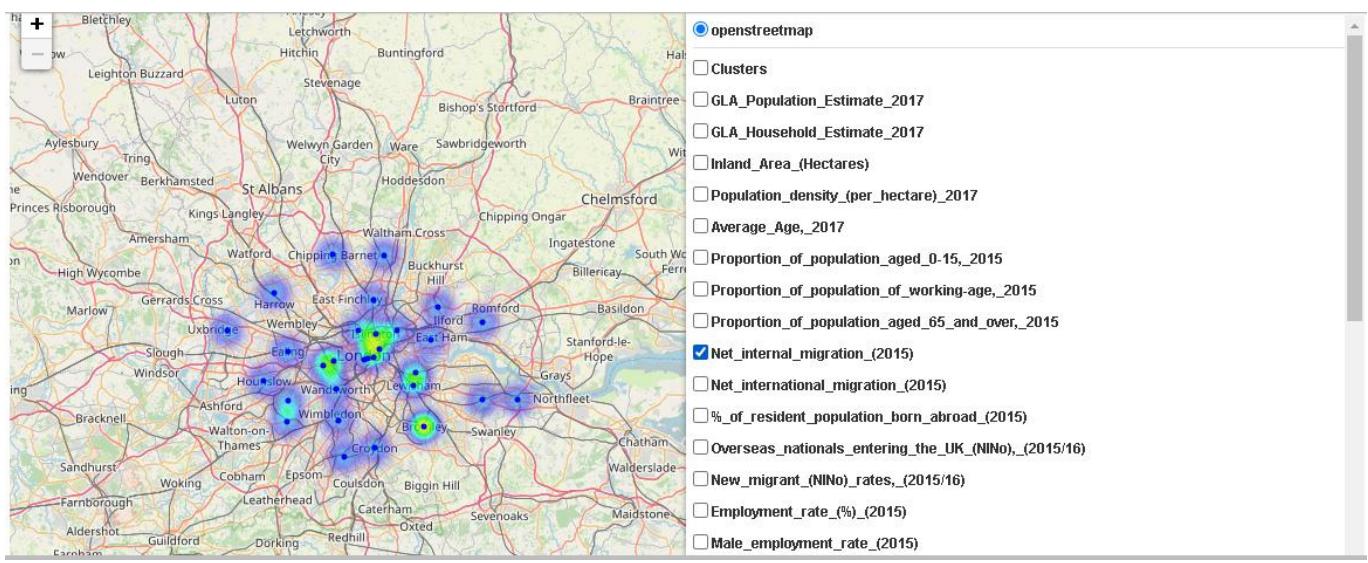
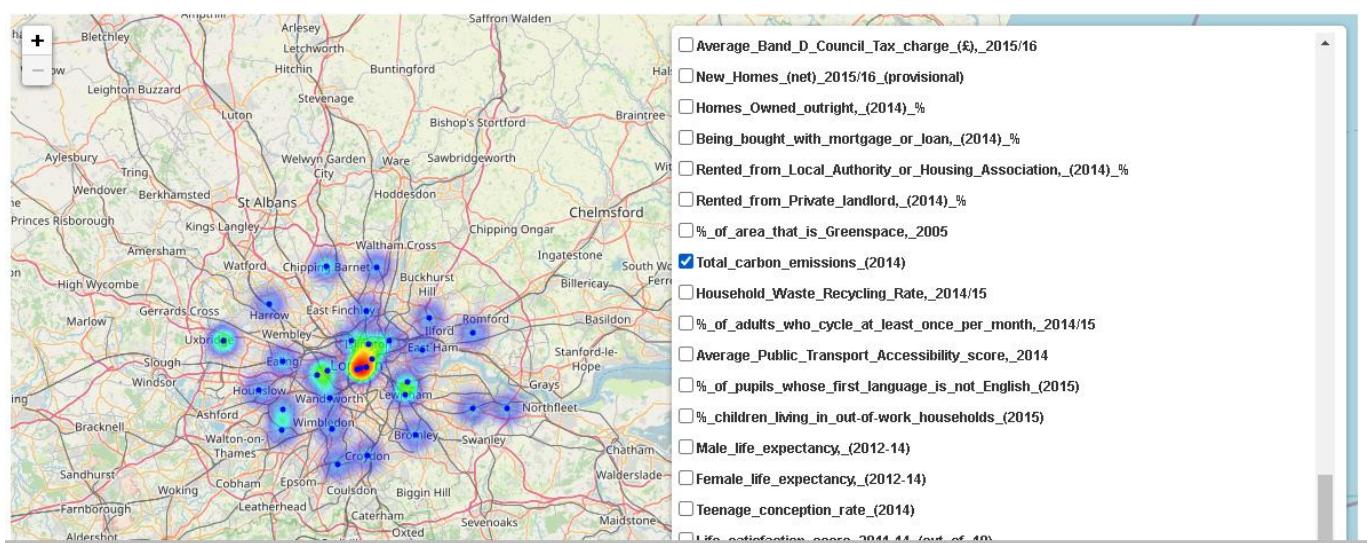
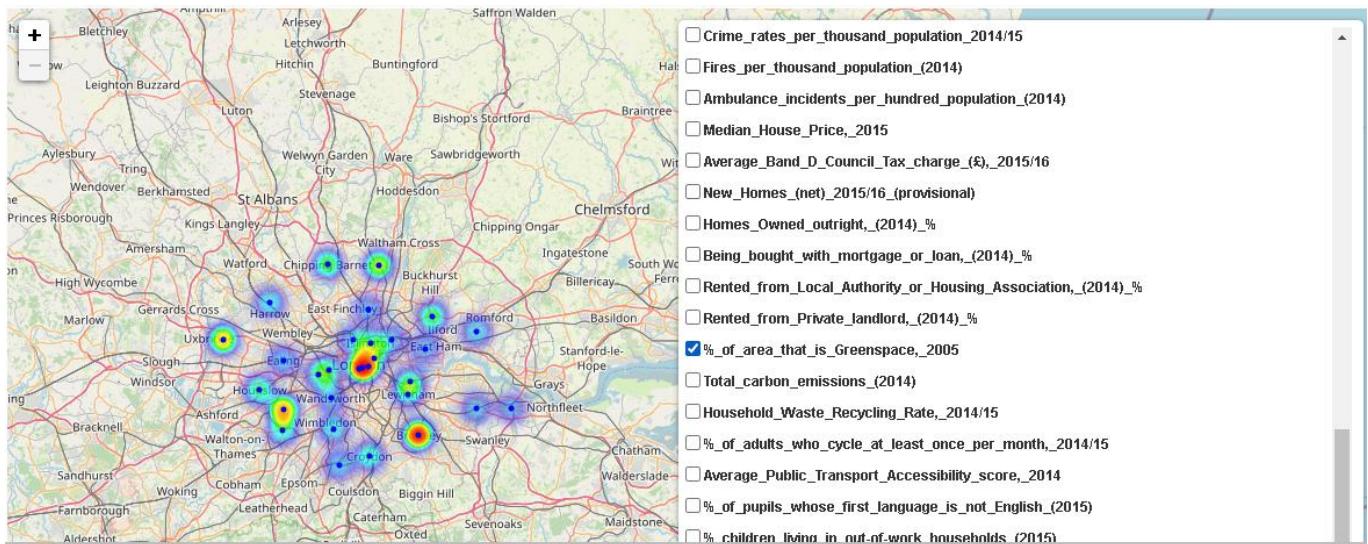
Barnet, Croydon, and Lambeth are the most populated boroughs with the most number of households as well. Barnet and Croydon, located in Northern and Southern London, respectively, the neighborhoods in these boroughs belong to Cluster no. 2 (Common Neighborhood Cluster) and Cluster no. 1 (Transportation Cluster). Apart from this, Lambeth is in central London; thus, its neighborhood is dominated by Cluster no. 7 (Master Cluster) and Cluster no. 5 (Sports and Fitness Cluster). With smaller inland-area and more population, Central London is the most densely populated among all boroughs:-



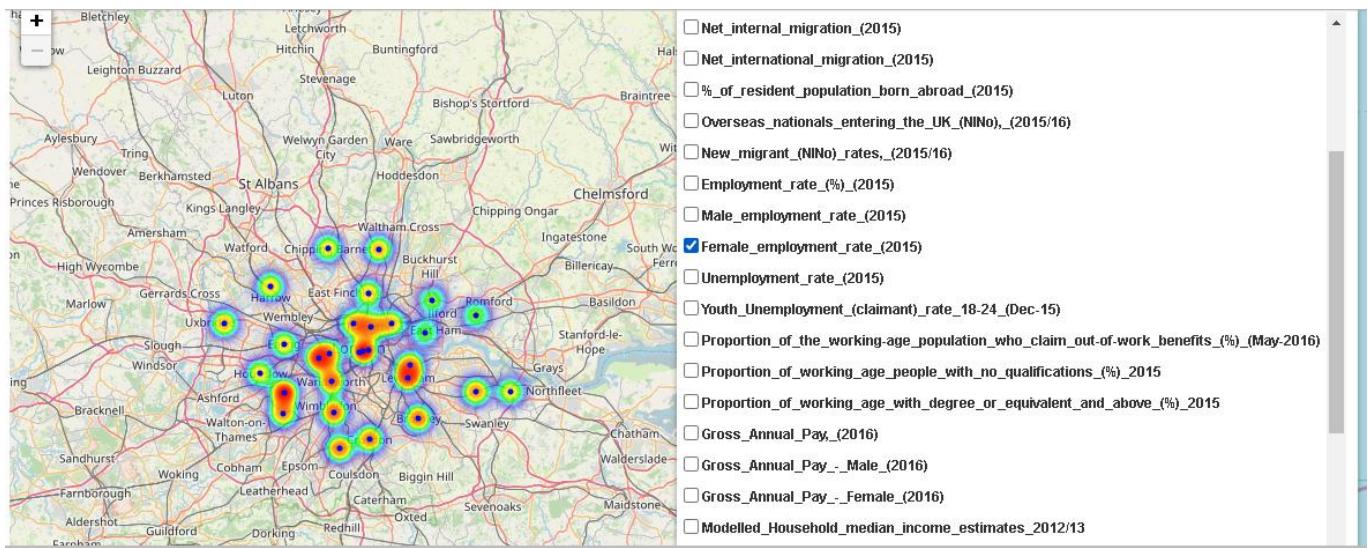
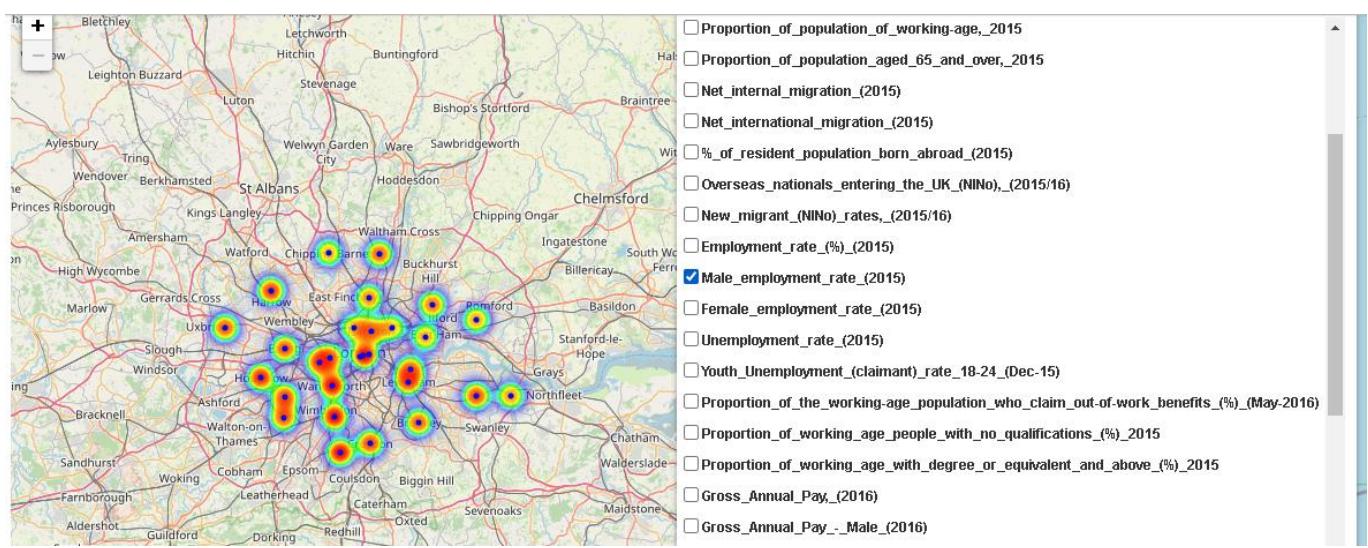
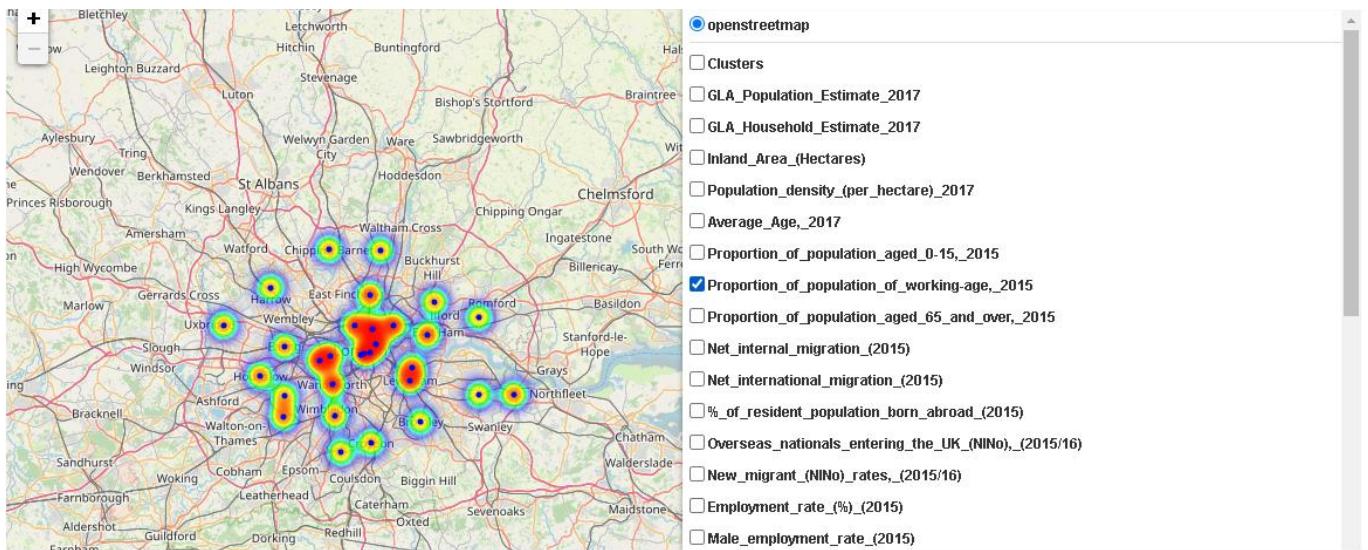


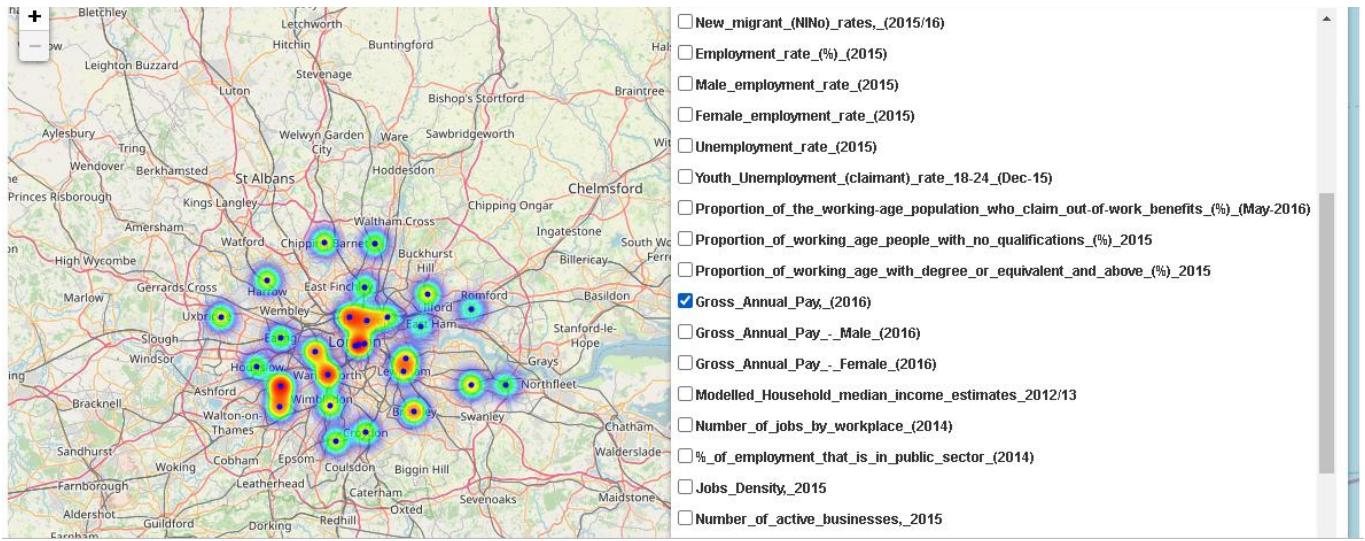
Bromley, along with Bexley and the City of London, have a significant chunk of the population, which is 65 years or above; therefore, the neighborhoods in these boroughs also include Cluster no. 6 (Pharmacy Cluster). Bromley, which is one of the 13 metropolitan centers of Greater London has the largest inland area in London with lesser population and households as compared to the other major centers like Croydon and Central Parts of London. It has the highest percentage of green space coverage and one of the least amount of carbon emission. As a result, Bromley sees most of the internal migration of the rest of the UK into London. Bromley and Bexley also saw the most number of houses bought outright (without and loan or mortgage):-



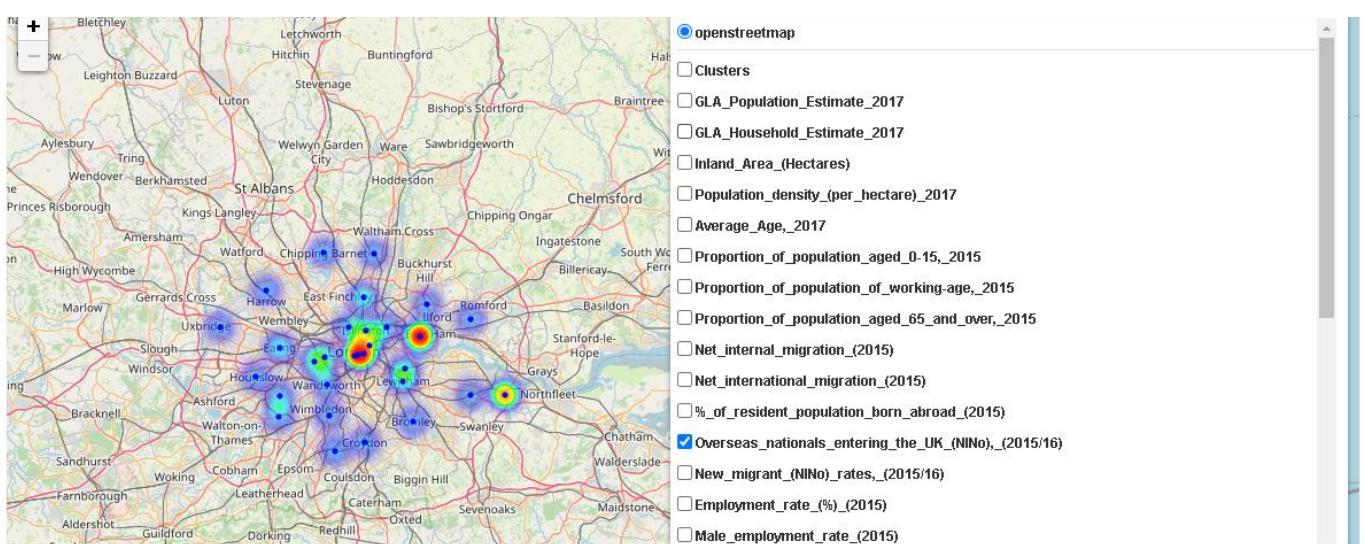
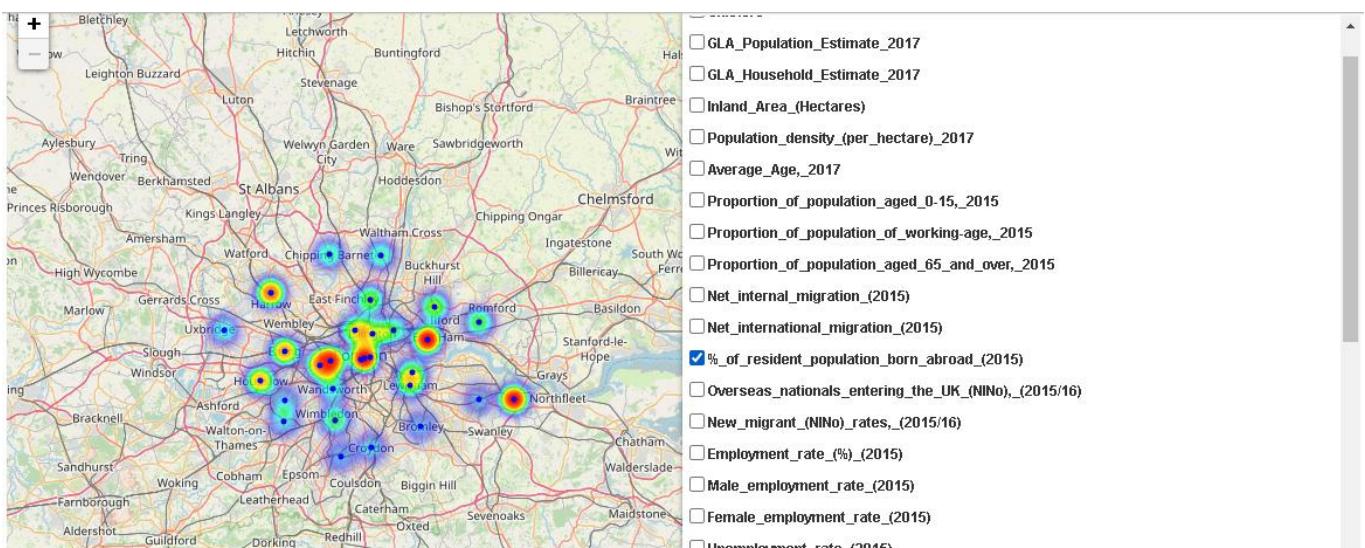


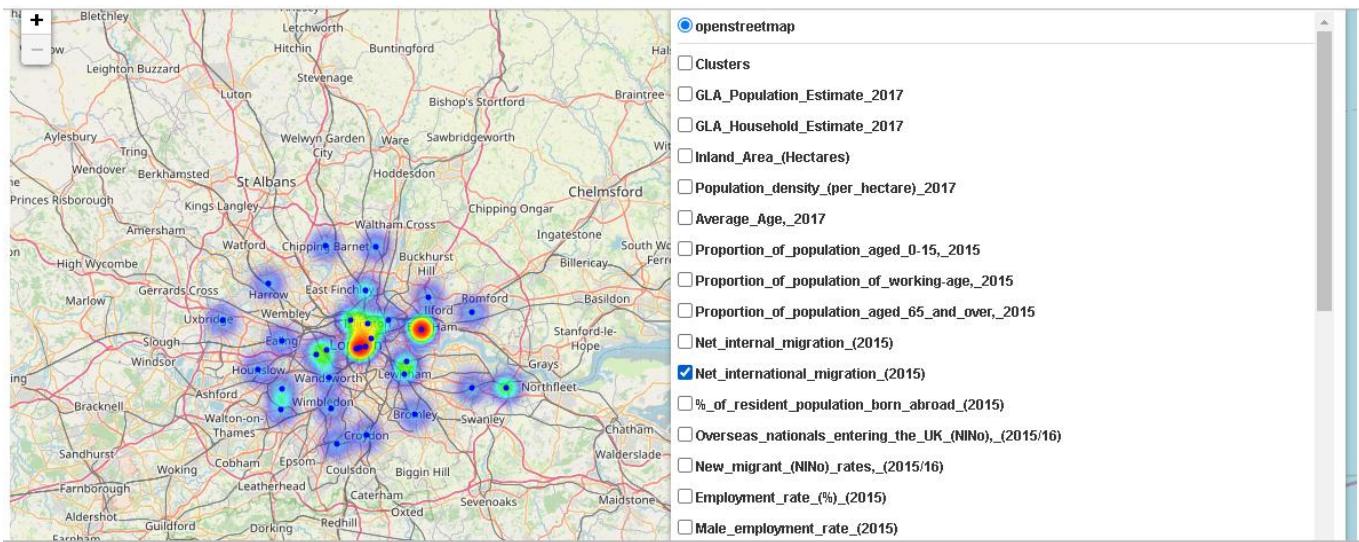
The majority of the working-age population resides in and around the areas of Central London, like the City of London, Islington, Westminster, Lambeth, Southwark, Camden, Hackney, Hammersmith, and Fulham etc. There are umpteen employment opportunities for both the genders in Central London. Apart from Central London, male employment opportunities are equitably available across London, with a significant chunk in the South-western part of London. On the other hand, Female employment is concentrated majorly at 'Richmond upon Thames' and 'Hammersmith & Fulham' boroughs. Thus female employment plays a crucial role in deciding Gross Annual Pay, and therefore 'Richmond upon Thames' has one of the highest Gross Annual Pay:-



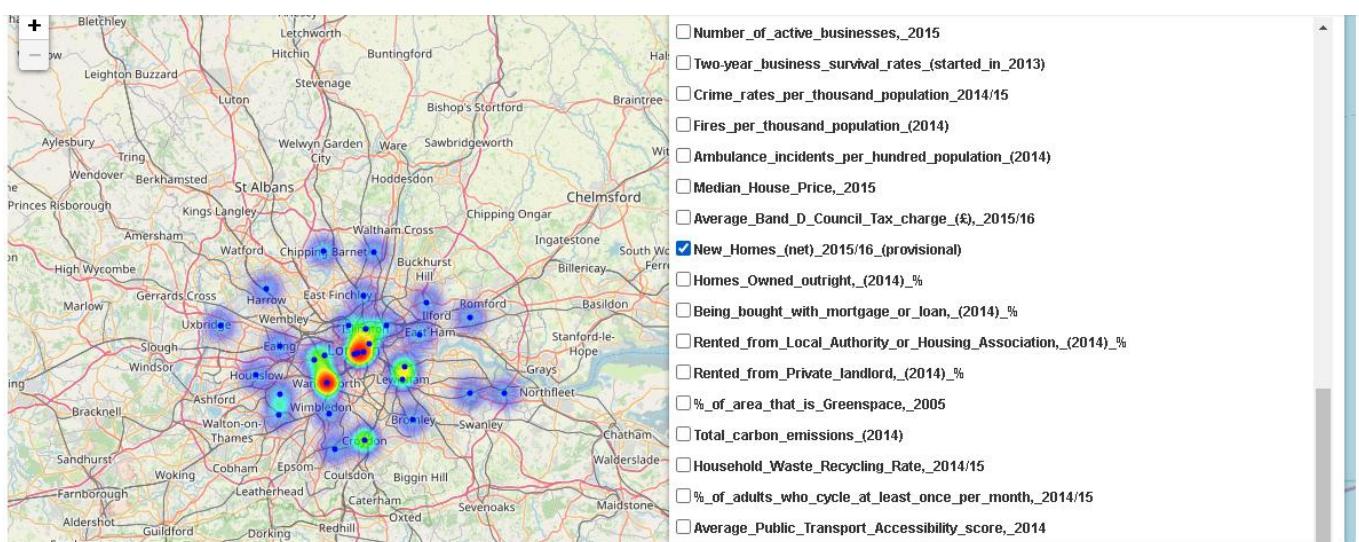
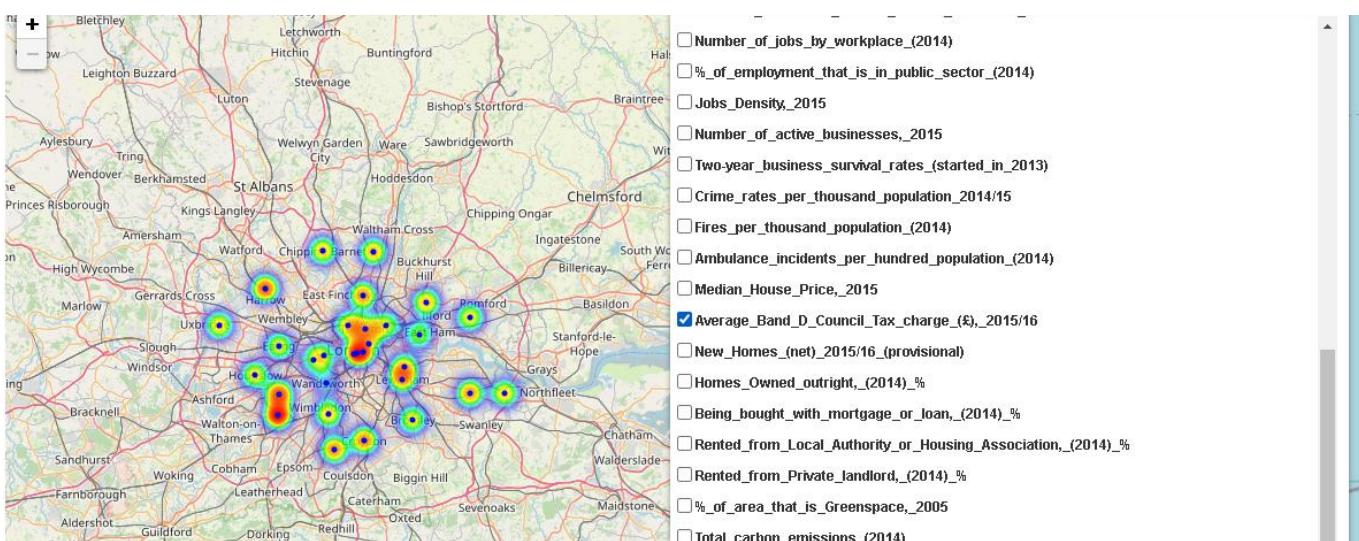


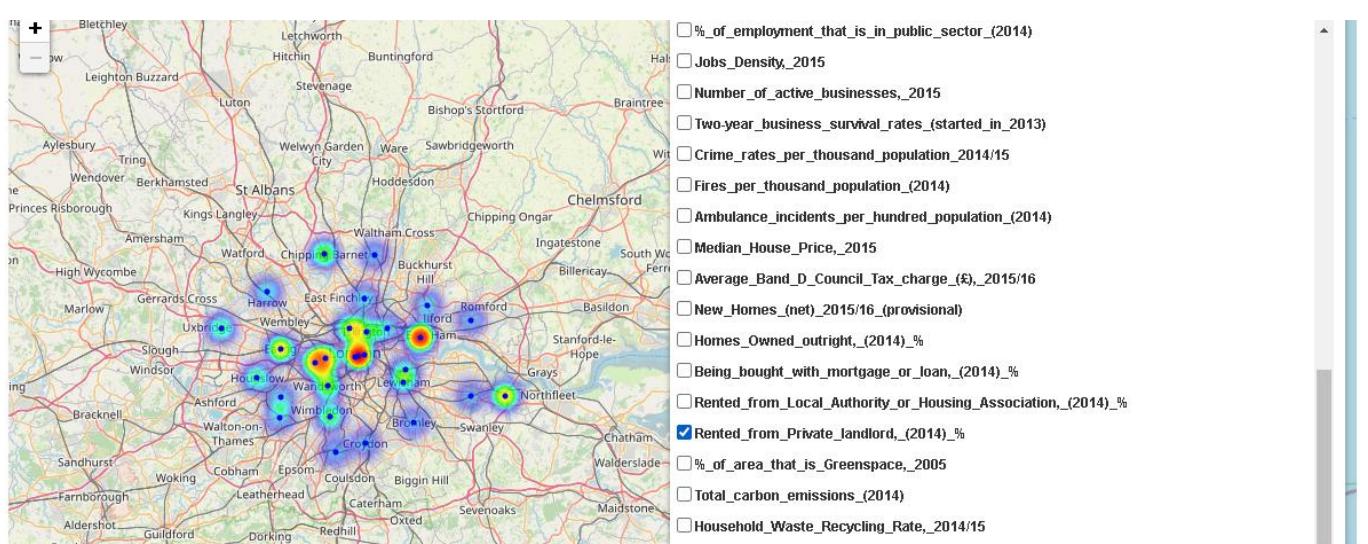
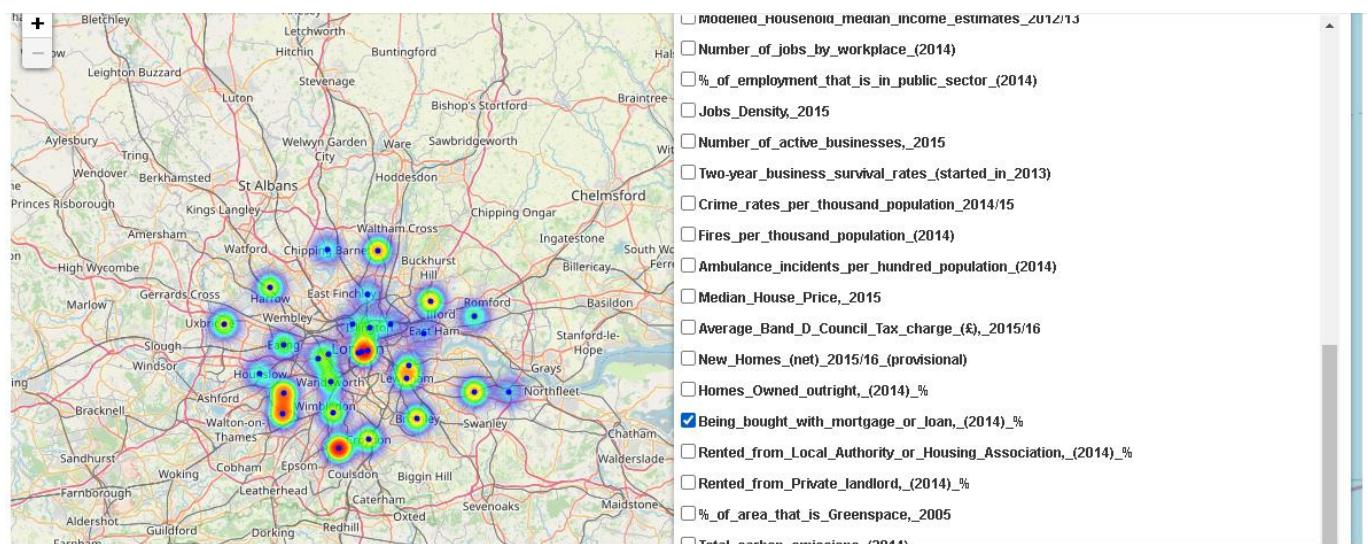
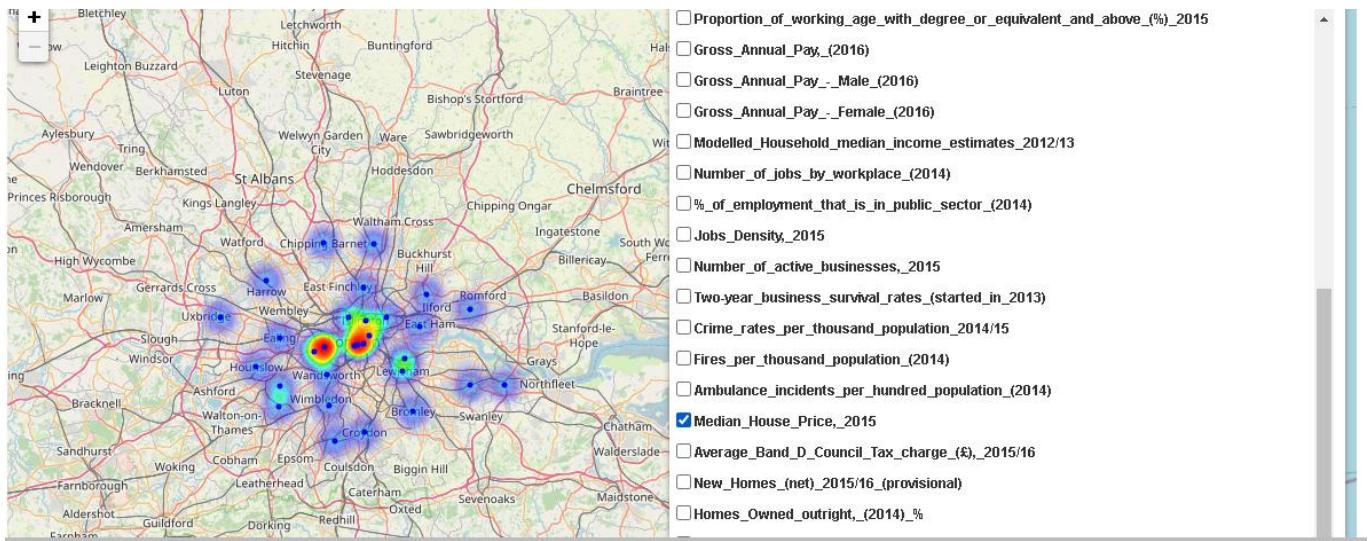
Westminster, Newham, Kensington & Chelsea, Harrow, and Brent have the most number of the resident population who are born abroad. Apart from Westminster, which is in central London and attracts international migration, Newham saw the most number of overseas nationals and international migration in the past few years. The neighborhood of Newham majorly contains venues from Cluster no. 1 (Transportation Cluster) and Cluster no. 10 (Grocery along with Cafe and Fastfood cluster):-

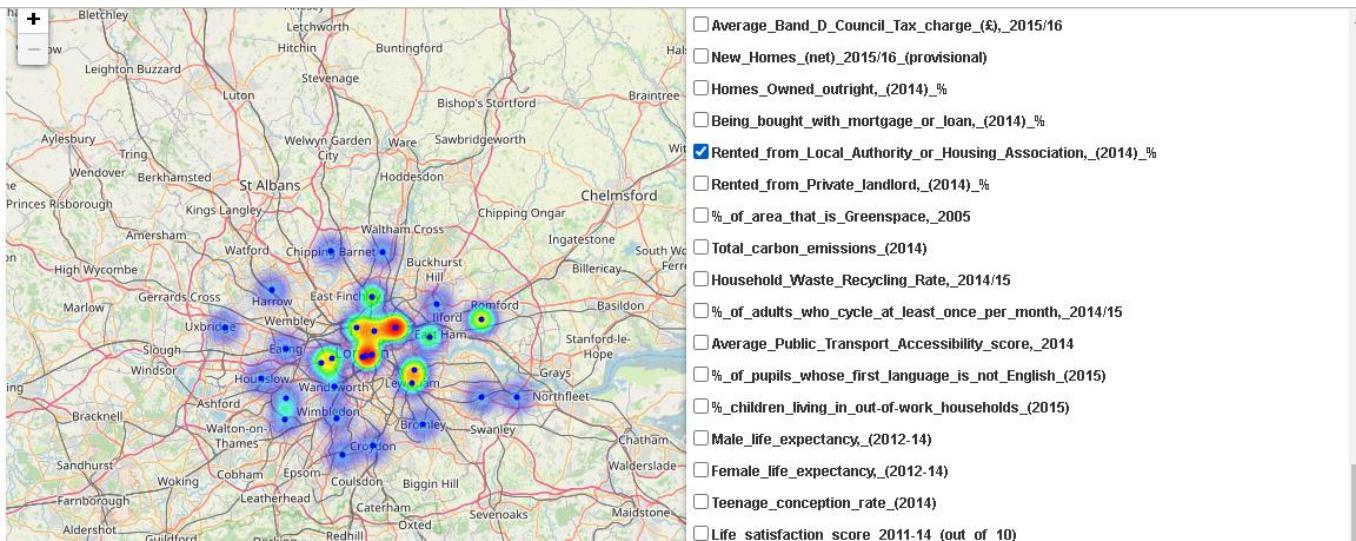




Wandsworth had the lowest average for the Band-D council tax charge, and thus it has the highest number of net New Homes. Kensington & Chelsea borough has the highest Median House Price since it is just around central London with Clusters no. 7 (Master Cluster), no. 5 (Sports & Fitness) and no. 10 (Grocery with Cafe and Fastfood) around it. On the other hand, Sutton in Southern London has the most number of houses bought on mortgage or loan. Newham has one of the most numbers of houses rented from private landlords. In comparison, Hackney has the most numbers of houses rented from Local Authority or Housing Associations:-



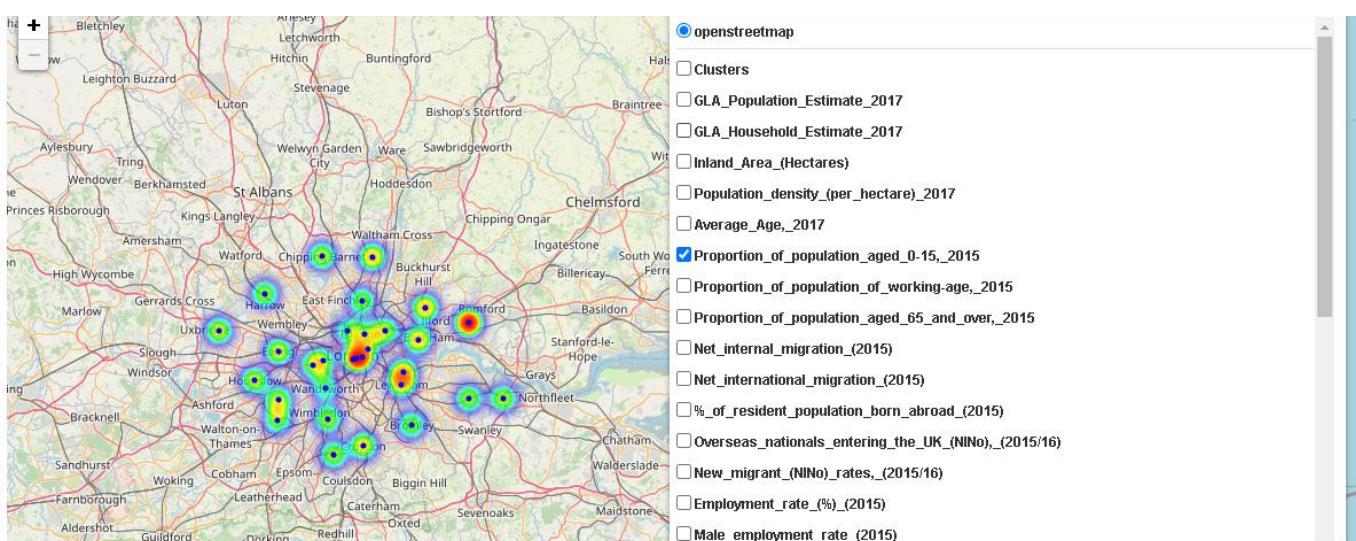


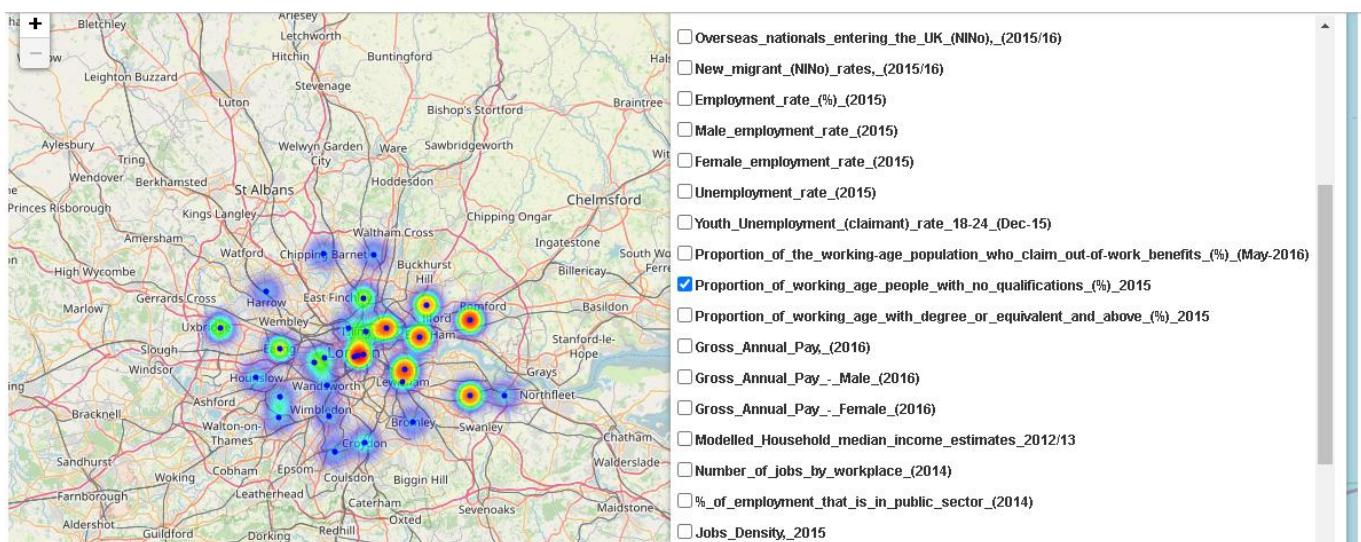
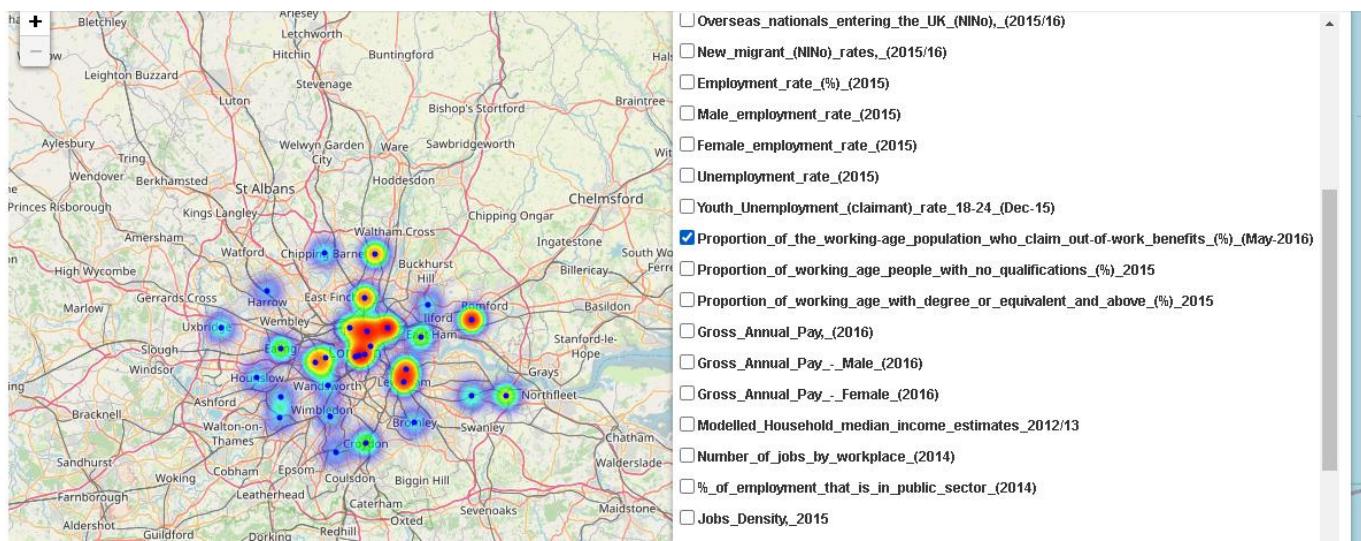
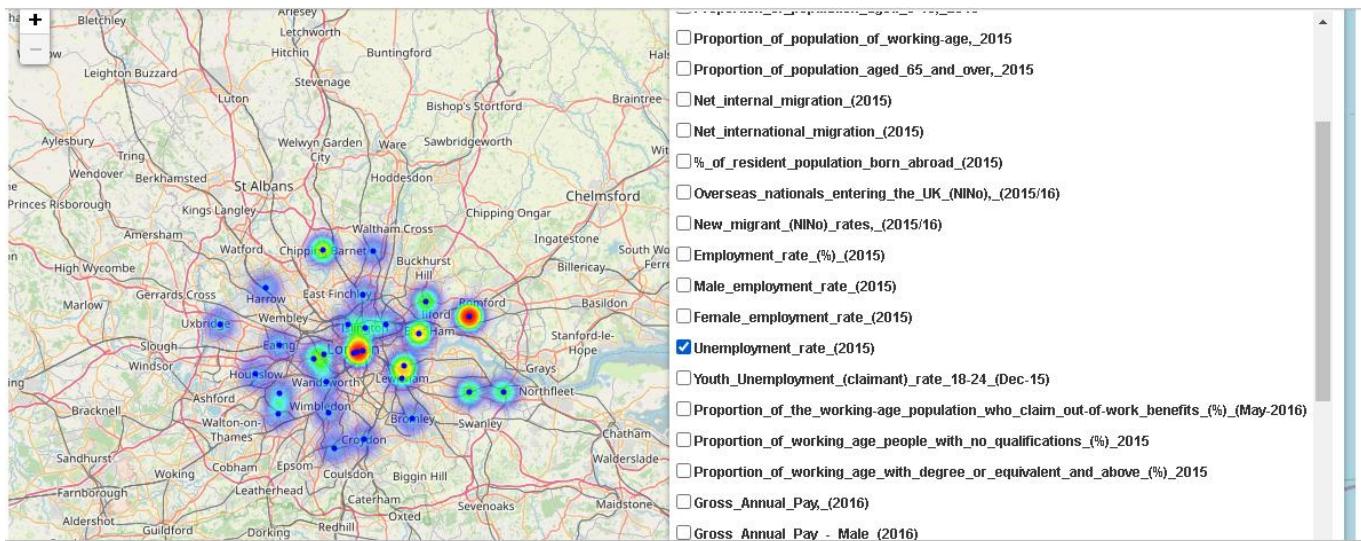


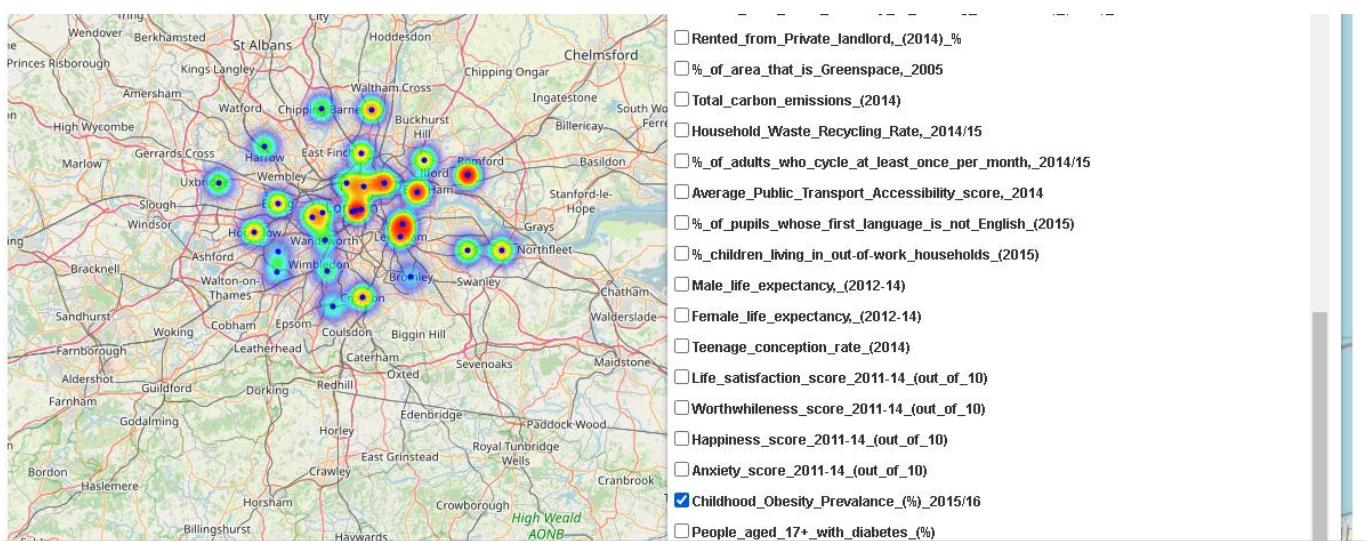
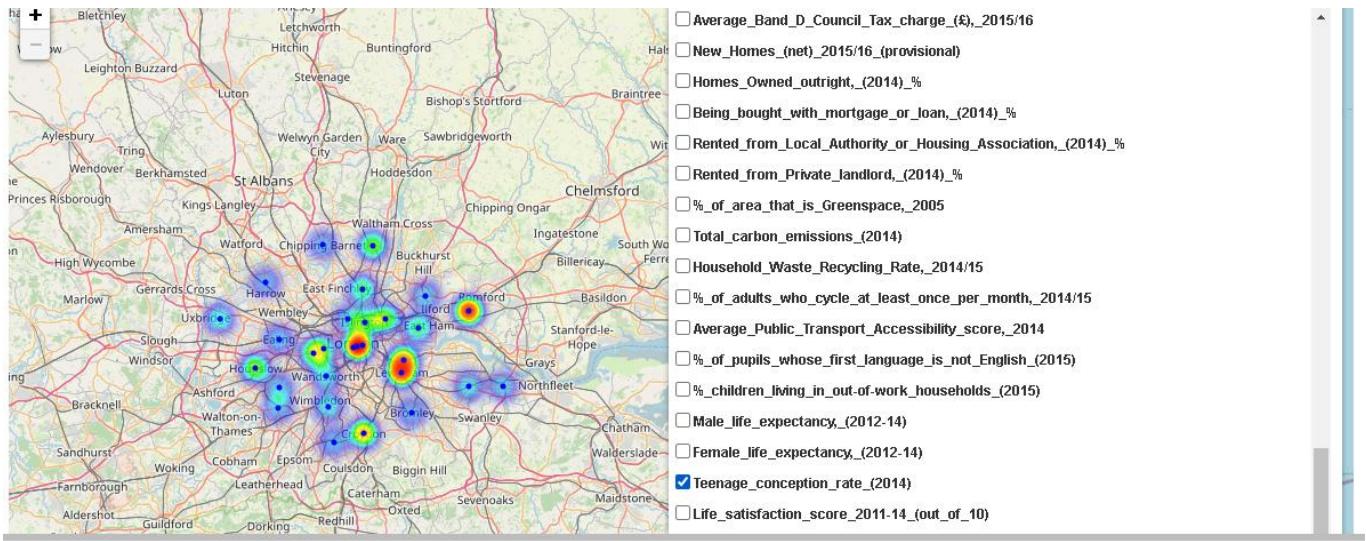
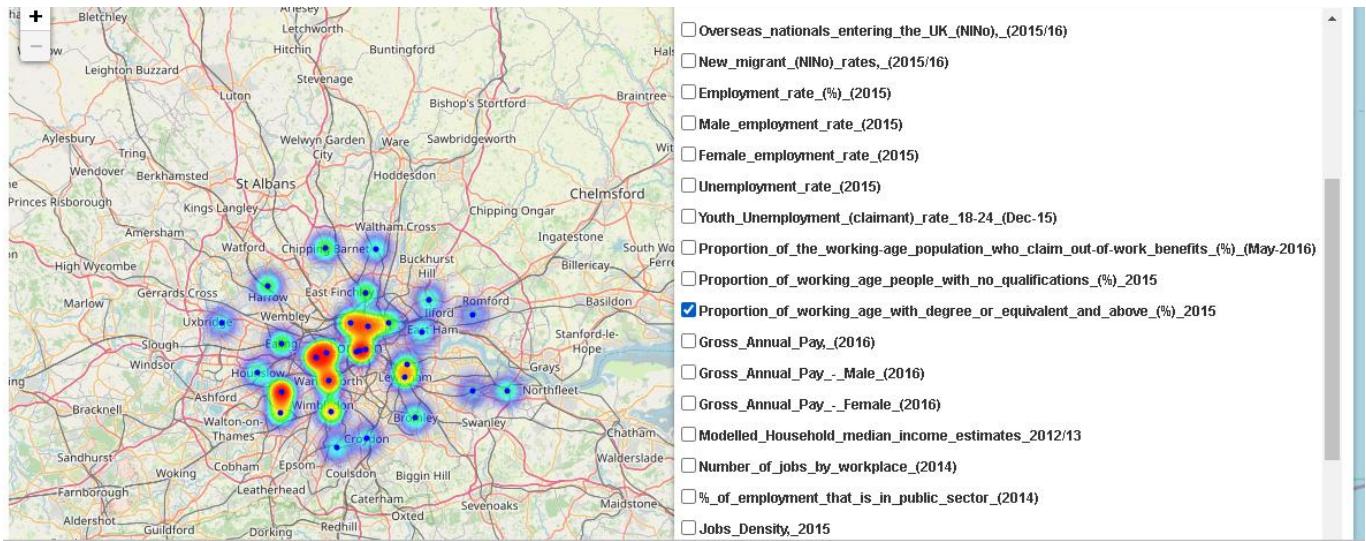
Problems of 'Barking and Dagenham' borough:-

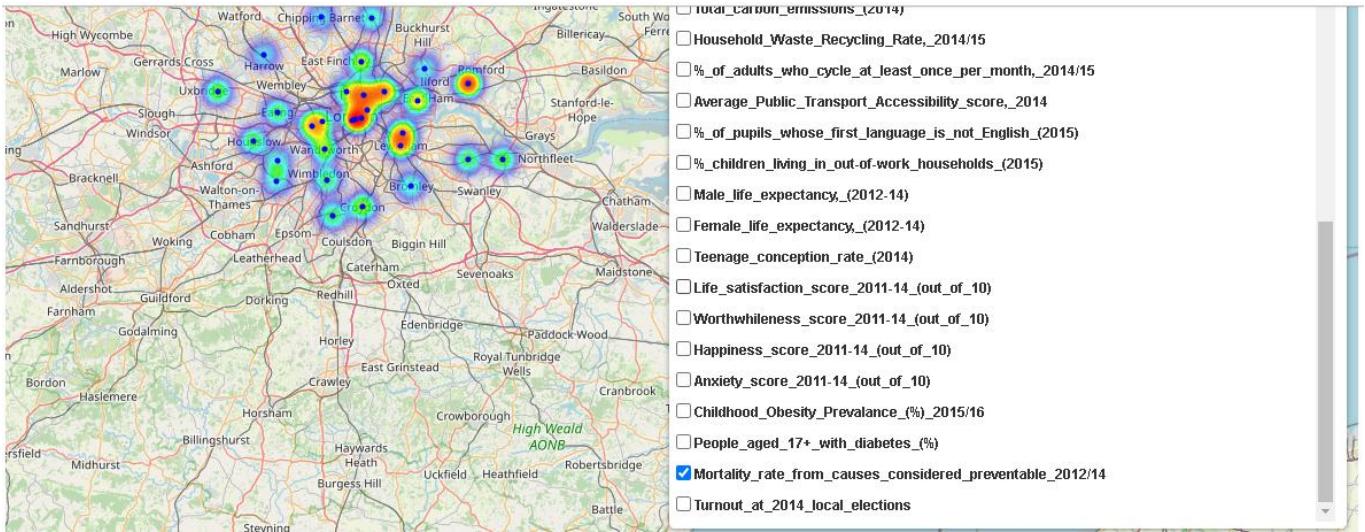
It is situated in the far eastern part of London and has the highest percentage share of the population aged between 0-15 years with One in every five jobs is in the public sector. there are following challenges with this borough:-

- Highest rate of unemployment, with every tenth person of the working-age population claiming out of work benefits.
- One of the highest percentages of the population, which is of working-age but has no qualification.
- One of the lowest proportion of the working-age population with a degree equivalent or above.
- Very high teenage conception rate.
- Very high childhood obesity prevalence.
- A high Mortality rate in which preventable causes cause deaths.









6. CONCLUSION:

As the Happiness-Score and Life-Satisfaction score, which are evenly distributed almost across the entire state, it is evident that London is a perfect right choice to move for any reason, be it studies, business, job, or just staying. Apart from that, it is not like there are problems with only one borough (*Barking and Dagenham*), and rest other boroughs have no problems at all. Other boroughs too have their problems, but they are distributed, whereas, in this particular borough, most of the problems are concentrated.

