Best London Boroughs to Live

Author: Umair Arif
Project: IBM Capstone
Date: 25th December 2020

1-Introduction

No London borough Is quite the same as another. There are 32 boroughs in total, they all have different attractions, architecture styles and atmospheres which makes them feel almost like a small cluster of different towns and cities.

We have been approached by a magazine, who wants to publish an article on the best boroughs of London to live in their upcoming edition. They have tasked us to determine the quality of life in each borough and the most ideal boroughs for outdoor activities, shopping destinations, entertainment and dining.

In this report, the data required to create the DataFrames on Python interface will be extracted from online databases such as London DataStore, Foursquare API, etc. once the dataframes are created, modelling packages such as Folium, matplotlib, etc. will be used to visualize the data in order to make it reader friendly, and finally followed by in depth analysis of the characteristics of each borough.

2-Methodology

This section illustrates the process the data goes through in order to be useful for the client. This part mainly consists of data extraction, data cleaning and data visualization.

2.1-Data Extraction

This process includes extracting the data from varies databases and convert it into Python friendly format, so that further analysis can be done on the data.

2.1.1 Quality of life data

The data required for quality of life analysis is extracted from Greater London Authority Datastore (Authority, 2019) Use the "Insert Citation" button to add citations to this document.
and it consists of happiness and transport accessibility score, % of greenspace, crimes per thousand and total carbon emissions in each borough. The extracted data is in excel format and can be imported on python interface using pandas dataframes to create the table as seen in Figure 1

	BoroughName	Crime rate per thousand	% of greenspace	Total carbon emission	Transport accessibility	Happiness score
0	Barking and Dagenham	83.359060	33.6	643.788244	2.970626	7.05
1	Barnet	62.738426	41.3	1415.107370	2.996701	7.37
2	Bexley	51.827942	31.7	974.621259	2.552134	7.21
3	Brent	78.801881	21.9	1175.073235	3.653713	7.22
4	Bromley	64.133577	57.8	1179.800587	2.779859	7.44

Figure 1: Shows how the quality of life data on Jupyter Notebook.

2.1.2 Geographical coordinates and population of each borough

The data needed for coordinates and population is extracted from Wikipedia page (Wikipedia, 2020) using web-scrapping package called Beautifulsoup. The tables created on python is as seen in Figure 2

	BoroughName	Population	Coordinates
0	Barking and Dagenham [note 1]	194,352	51°33′39″N 0°09′21″E / 51.5607°N 0.1557°E /...
1	Barnet	369,088	51°37′31″N 0°09′06″W / 51.6252°N 0.1517°W /...
2	Bexley	236,687	51°27′18″N 0°09′02″E / 51.4549°N 0.1505°E /...
3	Brent	317,264	51°33′32″N 0°16′54″W / 51.5588°N 0.2817°W /...
4	Bromley	317,899	51°24′14″N 0°01′11″E / 51.4039°N 0.0198°E /...

Figure 2: Shows the unfiltered data extracted from Wikipedia.

2.1.3 Data for venues

The data for venues is obtained by sending a request to Foursquare API. The data extracted is limited to top 50 venues in 500m radius of the center of each borough. The sample venues data in pandas data is as shown in Figure 3.

	BoroughName	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 School

Figure 3: Shows the sample of data obtained from Foursquare API.

2.2 Data cleaning and merging.

Once the data is extracted from the databases and converted into python friendly dataframes, the data undergoes cleaning process to remove unwanted columns and texts. Once the data is cleaned, the dataframes created in 2.1.1 and 2.1.2 are merged to create a single dataframes as shown in Figure 4.

	BoroughName	Crime rate per thousand	% of greenspace	Total carbon emission	Transport accessibility	Happiness score	Population	Latitude	Longitude
0	Barking and Dagenham	83.4	33.6	643.8	3.0	7.0	194352	51.5607	0.1557
1	Barnet	62.7	41.3	1415.1	3.0	7.4	369088	51.6252	-0.1517
2	Bexley	51.8	31.7	974.6	2.6	7.2	236687	51.4549	0.1505
3	Brent	78.8	21.9	1175.1	3.7	7.2	317264	51.5588	-0.2817
4	Bromley	64.1	57.8	1179.8	2.8	7.4	317899	51.4039	0.0198

Figure 4: Shows the cleaned and merged dataframe.

2.3 Data visualization

Once the raw data is converted into python friendly form, it is converted in visual forms for illustrative purposes and in order to make it easier for stakeholders to interpretate.

2.3.1 Quality of life data

The quality of life is plotted into horizontal bar graphs using matplotlib.pyplot package to plot the data against the boroughs. The following graphs are made to illustrate the data:

- Happiness score vs Broughs
- Transport accessibility vs Boroughs
- % of greenspace vs Boroughs
- Total carbon emission vs Boroughs
- Crimes per thousand vs Boroughs

2.3.2 Mapping London

The map of London with all its 32 boroughs labelled is created using Folium package This map will be further used in Section 2.4.1 to superimpose the results of K-means clustering

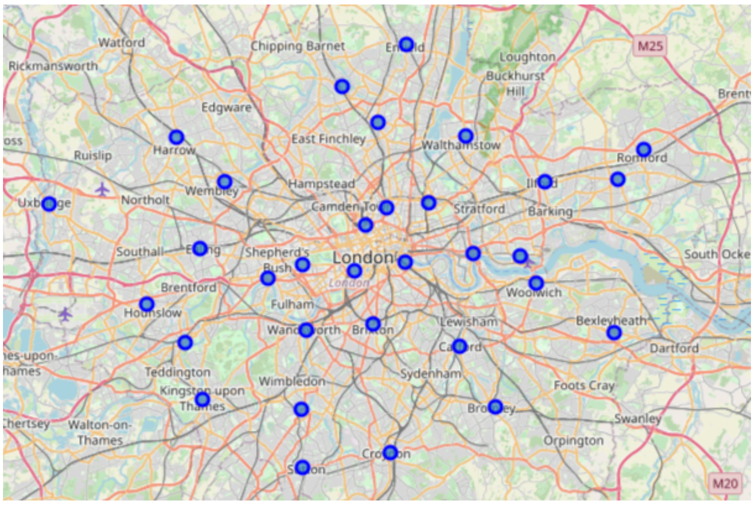


Figure 5: Shows the location of all 32 boroughs on the map of London.

2.4 K-mean clustering

Clustering is one of the most common exploratory data analysis technique used to get an intuition about the structure of the data. It can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different (Dabbura, 2018)

Kmeans algorithm is an iterative algorithm that tries to partition the dataset into *K*pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster (Dabbura, Towards data science, 2018).

- Specify number of clusters K.
- Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
- Keep iterating until there is no change to the centroids.
- Computer the sum of the squared distance between data points and all centroids.
- Assign each data point to the closest cluster (centroid).
- Compute the centroids for the cluster by taking the average of all data points that belongs to each cluster.

In this project, Kmeans is used to form clusters of boroughs with similar attributes, and to determine the boroughs ideal for outdoor activities, shopping destinations, nightlife and dinning.

2.4.1 using K-means of creating clusters of boroughs based on similarities

The k-means are create using sklearn.cluster, matplotlib.cm and matplotlib.colors. The data obtained is then converted to the a map using Folium package as shown in Figure 6.

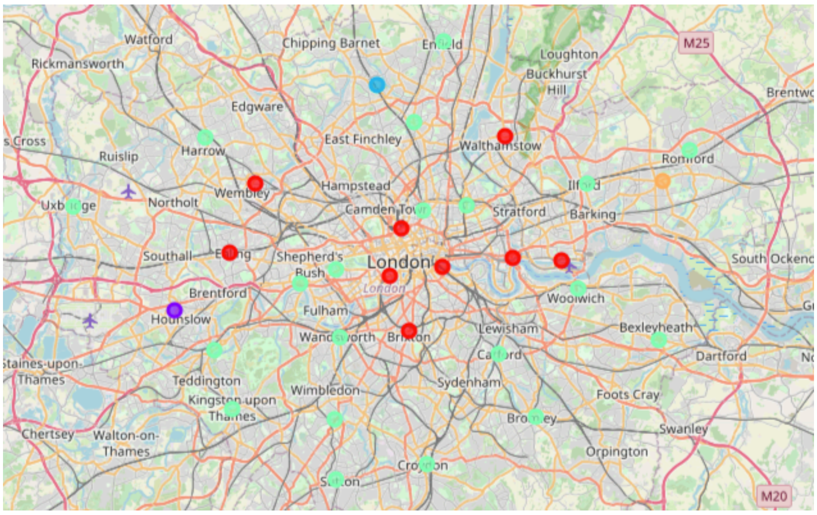


Figure 6: Clustering of London boroughs based on similarities.

2.4.2 Using K-clusters to determine top venue categories in London

The data obtained from Foursquare API is further process with sklearn.cluster to create five clusters. These clusters will classify boroughs on the basis of outdoor activities, shopping destinations, entertainment and dinning and boroughs which has a balance of all the venues.

3- Results and Discussion

In this section, the parameters used to analyze the quality of life and the data clusters created are discussed in depth.

3.1 Quality of life

3.1.1: Happiness Score

The happiness score takes the governance satisfaction, availability of public facilities, quality of educational facilities and availability of venues is taken into consideration. From the graph shown in Figure 7, Borough of Kensington and Chelsea performs the best in happiness score of 7.6, followed by Bromley and Kingston upon Thames with both scoring 7.4.

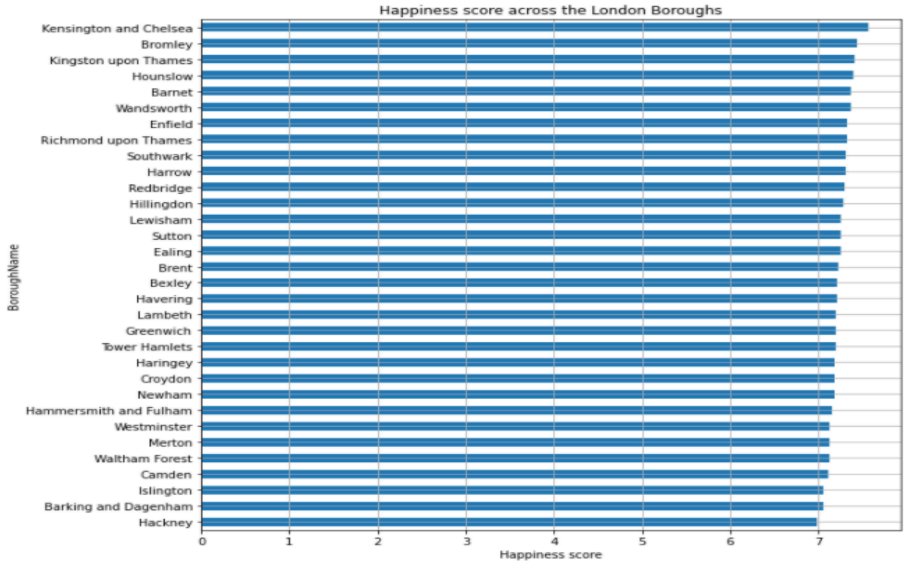


Figure 7: Shows the bar graph showing the happiness score across the London boroughs.

3.1.2: Transport Accessibility

Transport accessibility score is determined on the availability, frequency and quality of public transport infrastructure. The public transport fares and availability of cycling lanes is also taken into consideration. As seen in Figure 8, Borough of Westminster attains the highest score of 6.5, followed by Kensington and Chelsea and Islington scoring 5.8 and 5.7 respectively.

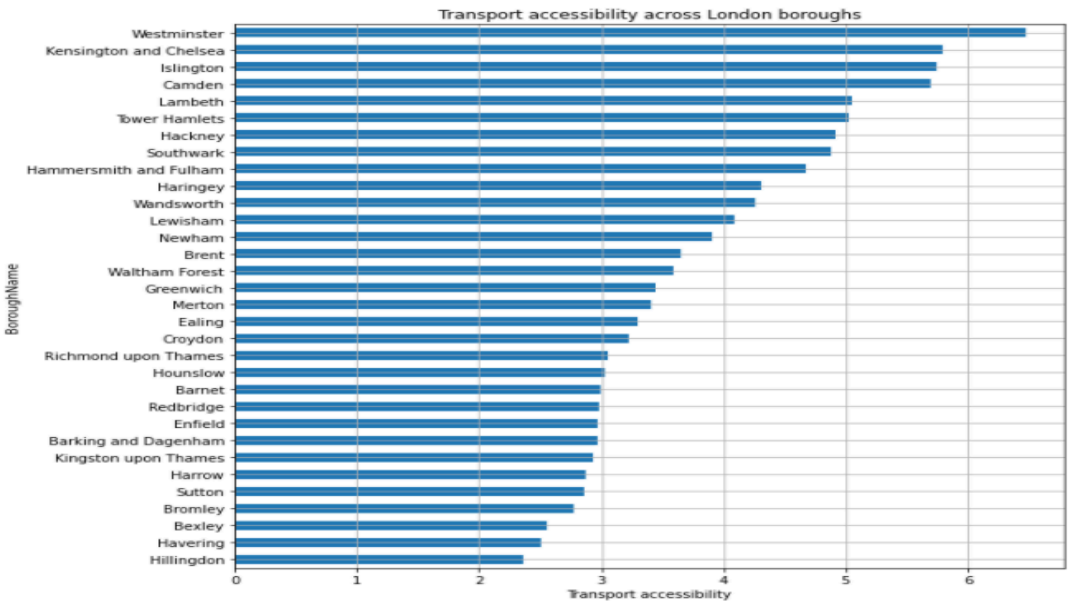


Figure 8: Shows the transport accessibility score of each borough.

3.1.3: Total Carbon Emission

Total carbon emission takes the pollution produced due to industries, commercial interties, residential units and transportation into consideration. From looking at Figure 9, it can be clearly seen that Borough of Westminster is the worst offender of all boroughs, producing almost 35% more carbon emissions than the second place, Borough of Hillingdon. The best performing borough is Barking and Dagenham, which only produces one fourth the emissions produced by borough of Westminster.

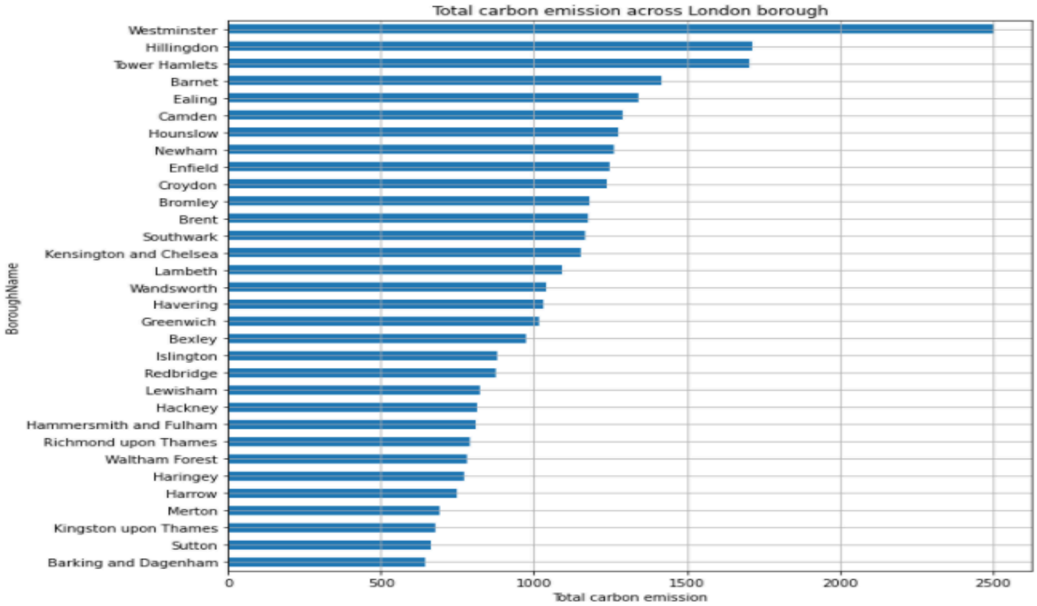


Figure 9: Shows the total carbon emission in each borough.

3.1.4: % of Greenspace

Percentage of greenspace takes account, the availability of parks and gardens, jogging and trekking tracks into consideration. In other words, the areas which is still left untouched. The borough of Havering tops this list (Figure 10) with approximately 60% of its entire area considered greenspace. The Borough of Islington has the least amount of greenspace available i.e. 13% of its entire area.

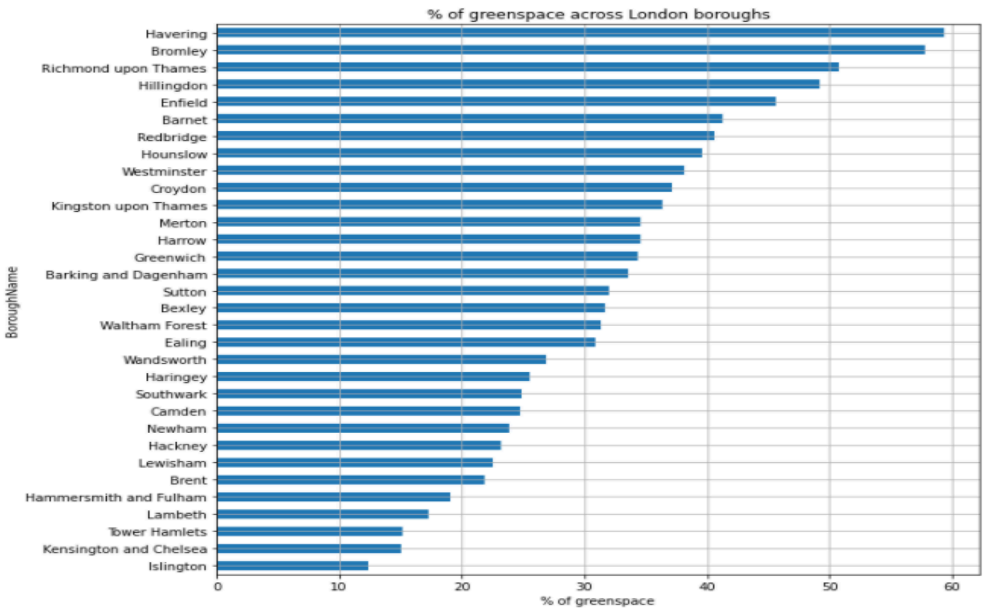


Figure 10: Shows the greenspace in each borough of its total area.

3.1.5: Crime Rate per Thousand

The crime rate is the amount of crimes reported in the boroughs per thousand people living in the borough. The Borough of Westminster is again the worst offender with reporting 213 crimes per thousand (Figure 11), twice as bad as second worst offender, Borough of Camden. The Borough of Harrow reported the least crimes per thousand with only 50 crimes per thousand.

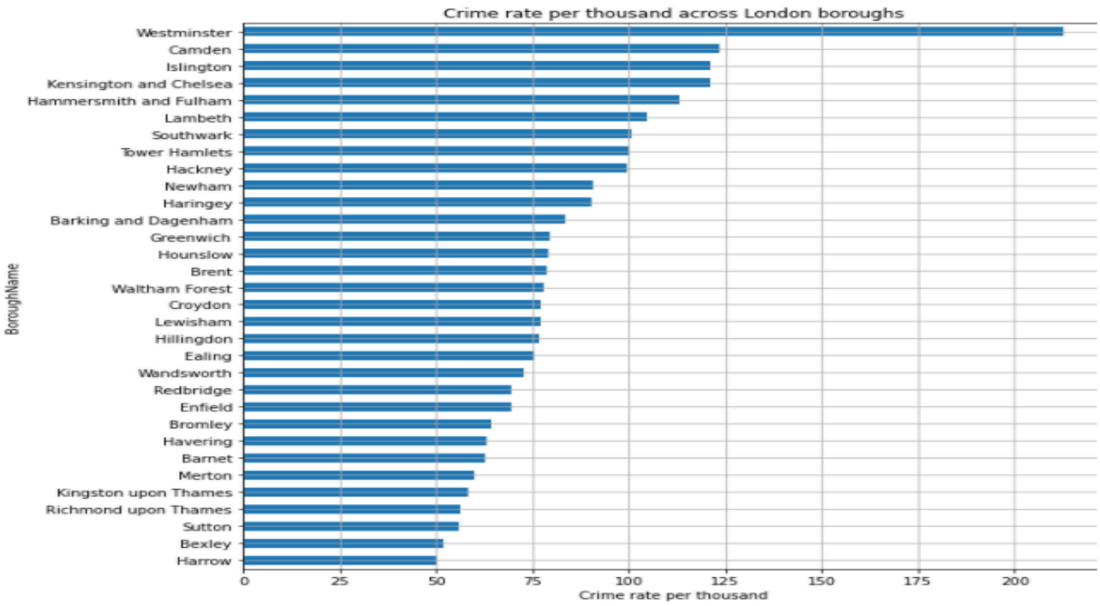
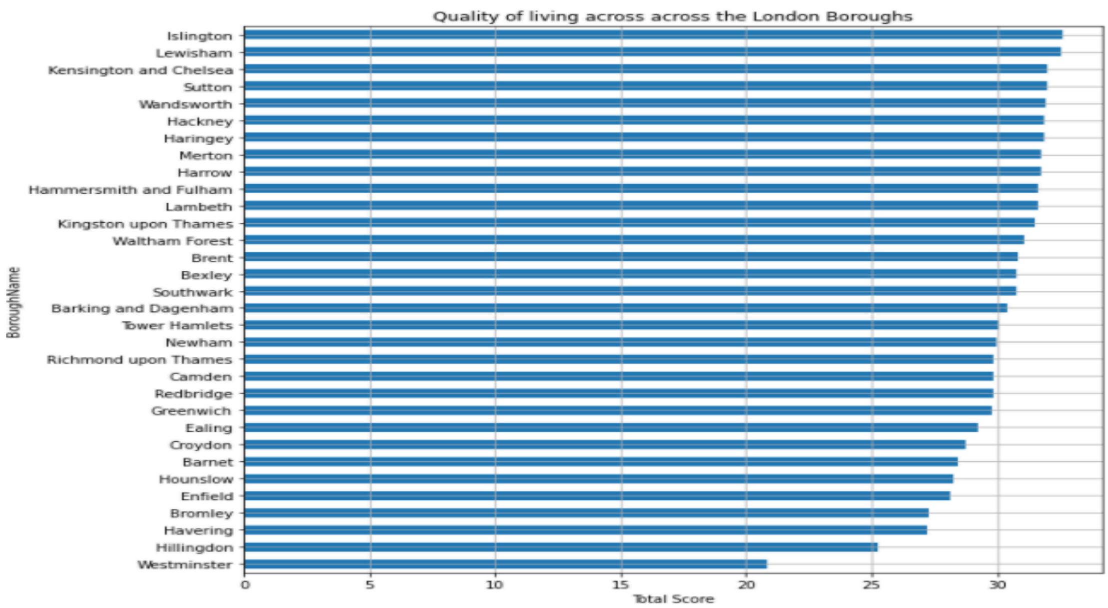


Figure 11: Shows the crime rate per thousand across the boroughs.

3.1.6: Total Score

The total score takes the happiness, transport accessibility, crime rate, % of greenspace and the pollution score into consideration. All the mentioned parameters are converted into score out of 10 and then combined to total score of 50 (Figure 12). The best performing borough using this criteria is Islington and Lewisham, both scoring 32.6. The worst scoring borough is Westminster, scoring 20.8.



Fire 12: Shows the overall performance of each borough.

3.2 Best boroughs in London for each amenities

This section discusses the results of K-mean clustering.
outdoor activities, shopping destinations, nightlife and dinning and boroughs which has a balance of all the venues.

3.2.1 Physical Activities

The most popular borough is Barking and Dagenham, with six of the top ten (Figure 13) venues being activities venues. It can also being seen that people in this borough usually prefer buses as their preferred mode of transport.

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
Pool	Bus Station	Gym / Fitness Center	Supermarket	Martial Arts School	Golf Course	Park	Dive Bar	Falafel Restaurant	Event Space

Figure 13: Shows the top ten venues in the borough of Barking and Dagenham.

3.2.2 Cluster 2: Shopping and dining

The most popular borough for shopping and dining is Bexley with seven of the top ten venue being shopping and dining destination (Figure 14).

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
Pub	Clothing Store	Fast Food Restaurant	Pharmacy	Supermarket	Coffee Shop	Portuguese Restaurant	Men's Store	Sandwich Place	Restaurant

Figure 13: Shows the top ten venues in the borough of Bexley.

3.2.3 Cluster 3: Entertainment

The most popular borough for entertainment is Barnet, with the six of the top ten venues stay open late night and are entertainment related.

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
Athletics & Sports	Construction & Landscaping	Bus Stop	Café	Dumpling Restaurant	Film Studio	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space

Figure 15: Shows the top ten venues in the borough of Barnet.

3.2.4 Cluster 4: Dining

The borough of Hounslow is the most popular destination for dinners, with seven of the top ten venues being different cuisines (Figure 16).

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
Bed & Breakfast	Pizza Place	Park	Café	Yoga Studio	Dumpling Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space

Figure 16: Shows the top ten venues in the borough of Hounslow.

3.2.5 Cluster 5: Balanced Boroughs

The borough of Brent is most balanced borough in all of London, where seven of the top ten venues fall under different venue categories (Figure 17) , indicating, it provides most options for the residents to choose from.

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
Coffee Shop	Hotel	Clothing Store	Sporting Goods Shop	Grocery Store	Sandwich Place	Bar	Restaurant	Stadium	Sports Bar

Figure 17: Shows the top ten venues in the borough of Brent.

4- Conclusion

To conclude with it, the borough of Islington and Lewisham scored the highest when it comes to quality of life, whereas the borough of Westminster performed the worst. Even though the borough of Islington came on top, Islington scored terribly when it comes to happiness score, crime rate and availability of greenspace. The only reason it came on top is due to its outstanding transportation accessibility and low carbon emissions. In the mean of all the five parameters used in calculating the quality of life is taken into consideration, the borough of Kensington and Chelsea comes on top, whose total score was seriously affected by the lack of green spaces.

When it comes to Clustering, we were able to classify the boroughs based on their similarities and also determine the most popular boroughs for physical activities, shopping, nightlife, dining and the most balanced boroughs.

5- References

Authority, G. L. (2019). *Great London Authority*. Retrieved from Great London Authority: https://data.london.gov.uk/dataset/london-borough-profiles

Dabbura, I. (2018, 09 17). *Towards data science*. Retrieved from Towards data science: https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a

Dabbura, I. (2018, 09 17). *Towards data science*. Retrieved from Towards data science: https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a

Wikipedia. (2020, 12 17). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/List_of_London_boroughs