# Comparing the demographics and facilities of London neighbourhoods
# 3rd January 2021

## 1. Introduction to the project: background and aim

London is one of the busiest cities in the world with over 8 million people calling it home. But do the demographics of the people who live in London, where they live and the facilities available to them, have distinct patterns? This project aims to investigate this.

We will look at both the demographics of the population and the types of venues in each borough of London to paint a picture of the people and their lives. This will be done in two parts

1. We will explore the boroughs by their demographic data. This will involve grouping the boroughs based on their demographic metrics and exploring their characteristics. This will provide some insight into the people living in each borough.
2. We will then use the Foursquare API to investigate the types of venues found in each area. This shows both practical facilities (shops, supermarkets, leisure centres) as well as sociable facilities (pubs, cafes, music venues). The analysis will include finding the most common categories of venues in each borough.

The types of venues in an area influence and are influenced by the people who live there. Therefore, with both of these analyses we can draw conclusions about the differences and similarities between boroughs.

The application of this could be to identify areas which could benefit from certain facilities based on their demographics. It could also lead to suggestions for government funding in certain regions. For example, does a borough with lots of young people but low happiness score need more youth clubs? Are there not enough supermarkets in a borough which has a high population density? How is the frequency of gambling shops linked to an area's demographics?

The audience for this could be a local government authority who wants to compare their facilities to boroughs with similar demographics. Or, on a city-wide level, it could also be useful to the government to identify regions which require more funding for certain facilities. It could also give an indication to people who want to open a business in London of where their audience lives and the other facilities in the area.

## 2. Data sources

The project will use two key datasets.

1. Demographic data by borough from Greater London Authority. This displays key population metrics such as population density, average age, % of resident population born abroad and happiness score. This will be used to characterise each London borough's population as described above.
   https://data.london.gov.uk/dataset/london-borough-profiles

2. Foursquare API data on venues in each borough. We will use the API to explore a radius around each borough and the different venues within this radius. These are split into categories such as restaurant, supermarket, pub etc. We can then calculate the proportion of facilities in different categories for each area. This is useful to understand the types of places that make up a neighbourhood.

## 3. Data analysis methodology and results

The following analysis makes use of the k-means clustering algorithm. This was chosen as it allows the data to be summarised so that conclusions can be drawn about each group.

### 3.1 Demographic data

We cluster the London boroughs into 5 groups taking into account the following characteristics:

- Population density (per hectare) 2017
- Average Age, 2017
- % of resident population born abroad (2015)
- Happiness score 2011-14 (out of 10)

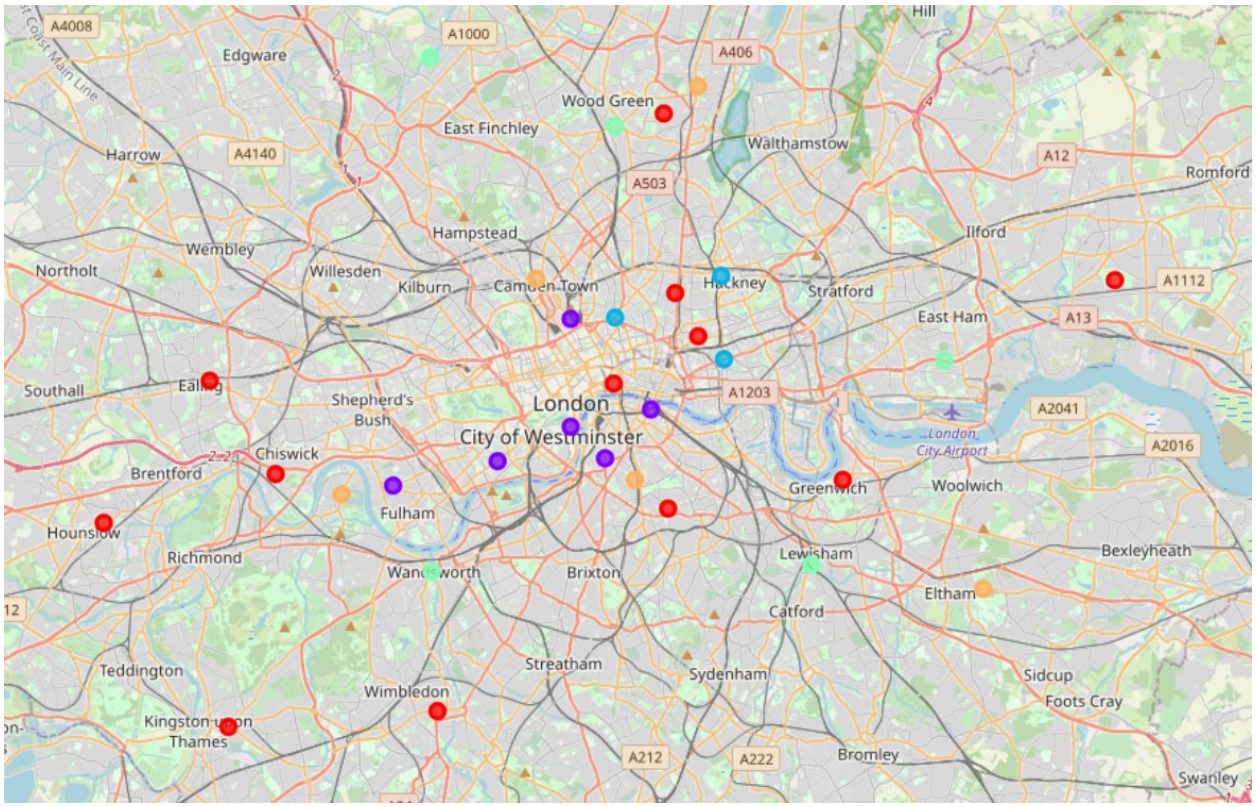
The 5 clusters have the following aggregated characteristics:

| Cluster | Population density (per hectare) 2017 | Average Age, 2017 | % of resident population born abroad (2015) | Happiness score 2011-14 (out of 10) |
|---|---|---|---|---|
| **0** | 51.052854 | 36.441667 | 37.216667 | 7.266667 |
| **1** | 116.615242 | 36.333333 | 42.816667 | 7.245000 |
| **2** | 151.099346 | 33.100000 | 37.000000 | 7.083333 |
| **3** | 86.106888 | 34.650000 | 42.083333 | 7.223333 |
| **4** | 29.037102 | 38.940000 | 20.280000 | 7.294000 |

Cluster 2 has the highest population density and lowest average age and happiness score. Cluster 4 has the lowest population density as well as the highest average age and happiness score.

We plot the results to see the geographic distribution of the clusters:

| Cluster 0 |
| --- |
| Cluster 1 |
| Cluster 2 |
| Cluster 3 |
| Cluster 4 |



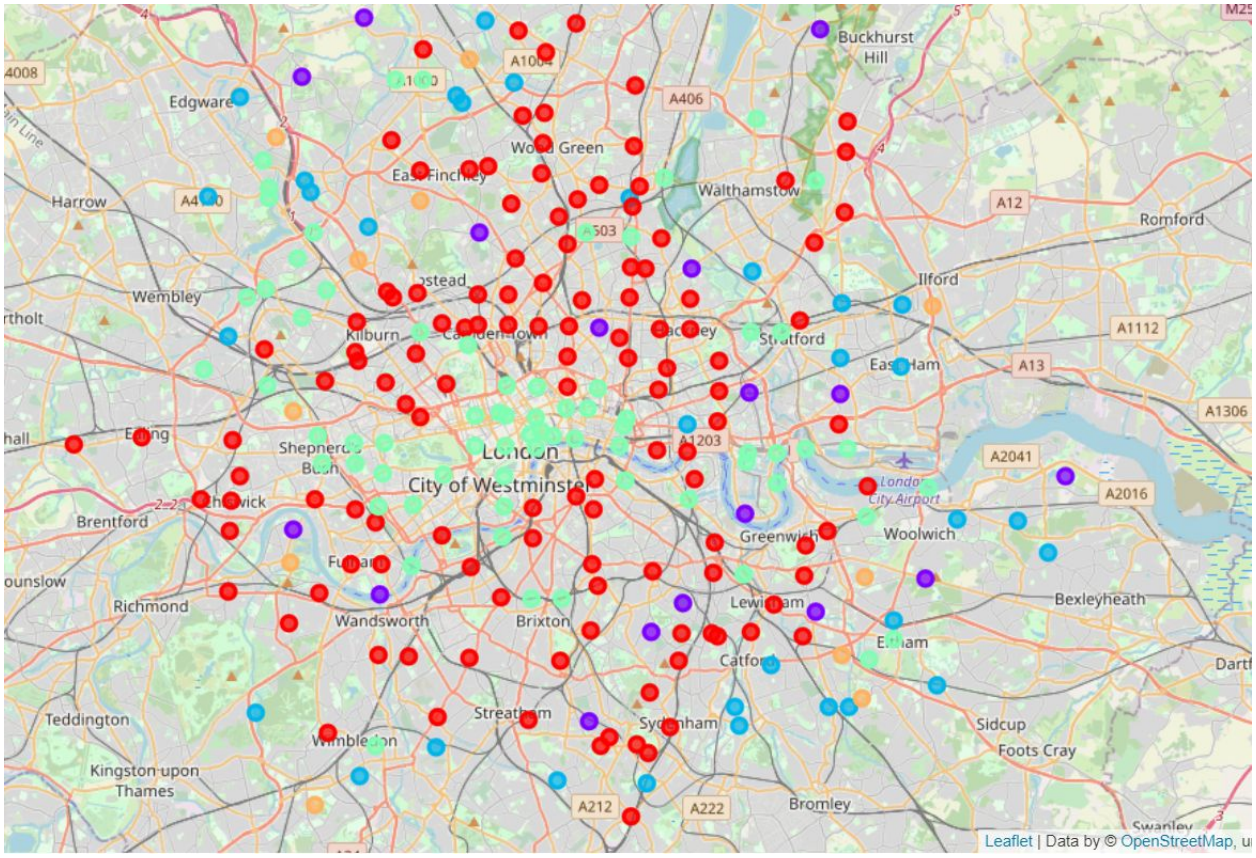

Cluster 2 boroughs are located in inner London, in the East. Cluster 4 boroughs are in the suburbs of London.

3.1 Demographic data

We now look at the types of venues in different areas of London. For each area we find the proportion of each type of venue and then group the areas into 5 clusters. We then plot the clusters on a map of London.

| **Legend** |
| --- |
| Cluster 0 |

| |
|---|
| Cluster 1 |
| Cluster 2 |
| Cluster 3 |
| Cluster 4 |




From looking at the location of each area, and the most common venues of each cluster, we make the following generalisations about each cluster:

- **Cluster 0** is the most common cluster with areas being found in a circle around central London. They stretch from inner London out to the suburbs. The most common venues are pubs, cafes and parks.
- **Cluster 1** areas are found amongst cluster 0 areas but are further out from central London. Pubs and cafes are common as well as outdoor venues such as zoos and golf courses.
- **Cluster 2** is found mostly on the edges of Greater London, representing suburban areas. The most common venue is a grocery store, as well as various nationality restaurants.
- **Cluster 3** is found in central London, along the river and in the North West of London. The areas in this cluster have a range of types of most common venues, from hotels to shopping to restaurants.

- **Cluster 4** is found also in the suburbs of London. The most common venues are parks and gyms.

## 4. Discussion and conclusion

From the results above we can see patterns in the demographic data of the people in London as well as in the types of venues present in each area.
Speaking about the general geographic areas:

- In Central London, the population density is high and there is a higher than average happiness score and average age. There are also a lot of different types of venues including shops, restaurants, zoos, hotels and ice cream shops. It may be the case that the venues are catered towards tourists and the areas are missing basic amenities such as supermarkets. This would require further analysis of specific venue types to compare across areas.
- In the area surrounding the centre of London, we have some areas with the highest population density and lowest happiness scores. Most Londoners would live in this section, and there would be large variation in the demographics of the people in each area. The most popular venues are pubs, cafes and parks which are expected for residential areas. Further analysis of the venues in the individual areas would provide insight into which areas have more facilities than others.
- In the outer areas of London we find the highest average age, lowest population density and highest happiness score. The types of venues are as you would expect for residential areas, grocery stores, restaurantes, as well as a lot of outdoor facilities like parks.

We can see that the centrality of a London location does somewhat correlate with the most common venues as well as the demographics of the population.

## 5. Future directions

Firstly, a more in depth analysis of the frequency of different venues in each area would lead to more insight on how well facilitated each area is in London. This could then be compared to the demographics of the population. For example, grouping the supermarkets, sports facilities or restaurants and comparing the density of them in each area. This may lead to suggestions for where areas are lacking facilities.

Secondly, a broader investigation into the demographics would also be insightful as this report only looks at four characteristics. This would add more detail to describe each area of London, and would surely lead to patterns of similarities between different areas in London.