

Data

Research commissioned by the Greater London Authority found 8,855 people slept rough in the capital between April 2018 and March 2019. However, these numbers are based on the estimates for the year. Furthermore, these numbers are reported by a source (Statista) that is not freely available.

We choose to use the official rough sleeping statistics for 2018, released on the 31st of January 2019 by the UK Ministry of Housing, Communities and Local Government. These statistics are based on counts and estimates carried out by Local Authorities, providing a snapshot figure of the number of people sleeping rough on any one night. As the figures reflect a snapshot, they differ in the scale to model-based estimates of other sources. Furthermore, some people rough sleep (bed down) on a single night and is never seen again (intermittent rough sleepers). Given that the true number is impossible to know, we rely on the official statistics of the UK Ministry of Housing, Communities and Local Government as shown in the link below:

<https://www.homeless.org.uk/sites/default/files/site-attachments/Homeless%20Link%20-%20analysis%20of%20rough%20sleeping%20statistics%20for%20England%202018.pdf>

In order to conduct our analysis on the socio-economic characteristics of the inner London boroughs, we get the data on variables such as population, unemployment, gross annual pay, crime rate, ambulance incidences and transport accessibility for each borough from the following source:

<https://data.london.gov.uk/dataset/london-borough-profiles/london-borough-profiles.csv>

There are 12 inner London boroughs and 20 outer London boroughs. “City of London” is not considered a borough and therefore removed from the data set. We calculate two rough sleep ratios for the boroughs: In the first ratio we divide the number of rough sleepers in inner London boroughs (as shown in the first link) by the corresponding hectare area of the borough. We scrape the borough areas from https://en.wikipedia.org/wiki/List_of_London_boroughs using the BeautifulSoup library. Furthermore, we use the population of the borough to calculate the second ratio (rough sleepers per 1000 inhabitant in each borough). The motivation behind the two ratios is to confirm that the boroughs Westminster and Camden have the highest rate of rough sleepers irrespective of which measure is used.

As a next step, we scrape the table from https://en.wikipedia.org/wiki/List_of_areas_of_London. This table has data on all London boroughs and areas, with the corresponding post codes. After some data cleaning/wrangling, we drop the rows other than the two boroughs we want to focus on: Camden and Westminster. These two boroughs have by far the highest rough sleeper rate among the inner London boroughs. As some locations ambiguously have more than one borough (for example, Brent and Camden), we conduct further checks to decide whether to assign the area to the Camden borough or remove it altogether. In order to make this check, we also downloaded (from <https://data.london.gov.uk>) the Mapping-template-london-ward-map-2013.csv. Although, we don't use the ward data directly, as the ward names often differ from the area names listed in Wikipedia. Nonetheless, the mapping table was a useful guide in removing the certain areas as the corresponding borough was not entirely in Camden or Westminster. These areas are Cricklewood and Tufnell Park.

The motivation for getting the postcodes is that we can get the latitude and longitude (coordinates) data for the areas of Camden and Westminster using the geolocator/Nominatim. Once that is done, the coordinates data are merged with the other data, and the main areas are marked in the London map.

Foursquare venues consists ten (top-level/first level) categories: 1. Arts and Entertainment, 2. College and University, 3. Event 4. Food, 5. Nightlife Spot, 6. Outdoors and Recreation, 7. Professional and Other Places, 8. Residence, 9. Shop and Service, 10. Travel and Transport. Using the coordinates based on the post codes, location and borough names, we queried the API by searching for the venues in Camden and Westminster boroughs. Our query was for the venues within 500 meter radius of each location with its corresponding latitude and longitudes. The resultant categorical data of detailed venues (127 unique categories) are converted to binary values through 'one-hot encoding', to use in the cluster analysis.