Battle of the Boroughs:

A Beer Garden-based segmentation of London

Pritpal Warner

July 2021

Data

There are six main sources of data as described below.

1. Borough geographic and population data from London Datastore

Description: file with columns listing area and population for each Borough.

Purpose: to inform exploratory descriptive analysis and Borough segmentation.

Source: London Datastore

Format: CSV

Link: https://data.london.gov.uk/download/land-area-and-population-density-ward-and-borough/77e9257d-ad9d-47aa-aeed-59a00741f301/housing-density-borough.csv

2. Borough demographic data from London Datastore

Description: file with columns listing the percentages of population 1) of working age (used as a proxy for pub-going age) 2) allowed by religion to drink alcohol. The data used was a subset of the source.

Purpose: to inform exploratory descriptive analysis and Borough segmentation.

Source: London Datastore

Format: CSV

Link: https://data.london.gov.uk/download/london-borough-profiles/c1693b82-68b1-44ee-beb2-3decf17dc1f8/london-borough-profiles.csv

3. List of London Boroughs and neighbourhoods from Wikipedia

Description: table of Boroughs and neighbourhoods in each Borough.

Purpose: to identify complete list of neighbourhoods for each Borough, for later use by Python Geocoder to return neighbourhood latitude and longitude.

Source: Wikipedia

Format: html table

Link: https://en.wikipedia.org/wiki/List_of_areas_of_London

4. Latitude and longitude data for each neighbourhood from Python Geocoder

Description: latitude and longitude coordinates for each neighbourhood.

Purpose: coordinates to be used in Foursquare API search/explore query to identify pubs with beer gardens in each neighbourhood.

Source: Python Geocoder

Format: Python lists

Link: !pip install geocoder

5.   Pubs with beer gardens in each neighbourhood and their coordinates

Description: list of pubs with beer gardens in each neighbourhood and their coordinates, using categoryId = 4bf58dd8d48988d11b941735 ('pub') and query='beer garden' to return the required information.

Purpose: to identify beer gardens, informing exploratory descriptive analysis and Borough segmentation through clustering.

Source: Foursquare search/explore query API

Format: .json file

Link: https://api.foursquare.com/v2/venues/explore?client_id={} &client_secret={}&ll={},{}&v={}&categoryId=4bf58dd8d48988d11b941735&query=beer garden&limit=100'.format(lat, long)

6.   Borough boundary polygons

Description: file containing latitude and longitude coordinates specifying the geographical boundaries of each Borough.

Purpose: to enable 1) correct assignment of pubs to Boroughs during processing of the Foursquare query results, 2) visualisation of clusters of Boroughs after clustering.

Source: https://skgrange.github.io/data.html

Format: .json file

Link: https://skgrange.github.io/www/data/london_boroughs.json