Data

The data I will be using for this project are listed below:

* Foursquare API data
* Table containing list of London areas
* London crime rates by borough
* London unemployment rates by borough
* Average house prices in London by borough

I will describe the content of each data, its sources and how it will be used.

Table containing list of London areas

This table will be scraped from the following Wikipedia page: <https://en.wikipedia.org/wiki/List_of_areas_of_London>

This will be used as it is a useful table sorting different areas in London by borough.

Foursquare API data

Foursquare API data will be used to obtain the list of venues and categories for each London areas listed in the table scraped from above Wikipedia page. The ‘explore’ endpoint will be used to obtain the list of venues. After this, the list would be reorganized so that it will become a table showing frequencies of different categories of venues.

London crime rates by borough

This is a table showing crime occurrences in London by borough. Obtained from the following website: <https://www.finder.com/uk/london-crime-statistics>

The data from this source is slightly not up to date, with the latest statistics coming from the year 2018/19, but it was the only source I could find that sorted crime occurrences in London by borough. This table will be scraped by beautiful soup, and crime rates for each borough will be calculated by dividing it with borough population.

London unemployment rates by borough

This is a excel file containing employment related statistics by borough. This file is obtained from the following website: <https://data.london.gov.uk/dataset/economic-activity-rate-employment-rate-and-unemployment-rate-ethnic-group-national>

The specific sheet containing unemployment rates by borough will be saved as csv file, and I will import it as pandas dataframe. Like the crime occurrence data, this data will also be slightly not up to date, with the latest stats coming from the year 2018/2019. Necessary modifications would be applied to make it suitable for clustering.

Average house prices in London

The following page contains average house prices in London by borough: <https://www.statista.com/statistics/1029250/average-house-prices-in-london-united-kingdom-by-borough/>

I will incorporate the figures from the graph in this link to the dataframe after clustering. Average house prices will not be used for the clustering process, but it is important data necessary for discussion and making a conclusion about which boroughs are potentially worth investing in.