

Capstone Project – The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera.

Problem: Where to start the search when relocating from central London to the Home Counties

Data

Following is the data required to solve the above problem and how each type will be used.

- **Postcode Districts in London and Home Counties and respective coordinates.** This data looks very similar in format to the data used for the Toronto project, i.e., postcode (e.g. KT2), Name, County/London and longitude and latitude coordinates. The data comes from a GitHub csv file, which is based on Ordnance Survey data. This data will be the baseline to connect postcode districts, location coordinates and wider area location (London/County).
- **London coordinates** from geocoder to get the Folium map for London.
- **Foursquare data** of venues within 500 yards of each postcode district. The analysis will be very similar to the clustering analysis in the Coursera course to find the top/most common venues in each borough/postcode district. We will use this data to cluster postcode districts and find similar ones.
- **House price data by county.** The Office for National Statistics provides a downloadable file of quarterly average house prices by county - from 1995. To aid this analysis we provide the average house prices for each home county and a simple prediction of future house prices, should these continue to grow at recent rates.