**Abstract**

**Introduction**

In a large metropolitan area such as London there are a wide variety are neighbourhoods and boroughs each with unique selection of restaurants, cultural sites, night life est. and it is my hope to explore the relationship of the neighbourhood profiles and identify if any if any relationships exist between these profiles and residential housing prices or crime rates, within the city.

This is of interest to me as a new resident of London and am looking to purchase property in the city. I want to find a suitable neighbourhood for myself that satisfies three criteria a happy medium of reasonably priced areas, that a safe and offer a strong night life entertainment. To determine the ideal location for myself I will explore the relationships between these three variables within the neighbourhoods within London. I will determine the success for this project if I am able to identify the ideal postcode for myself to purchase a property within the city of London that best matches my requirements.

I find myself needing to utilise this capstone project to accomplish this task as I have found no quantifiable means of determining a good fit neighbourhoods for a new resident and most information is very vague and offers little in the way of making an informed decision on the topic.

To be able to answer the question I started by breaking down the question of “What neighbourhood best satisfies my desired profile of safe, reasonably priced and safe”. From this statement I determined I needed a few key pieces of information to start with:

1. A way to identify neighbourhoods at a granular level
2. A way to geolocate these neighbourhoods
3. Neighbourhood demographic data to help with clustering analysis
4. Data base of venues that can be grouped by geolocation which included information on nightlife at a minimum
5. Data repository of housing prices sortable by a unique neighbourhood identifies
6. Data repository of crime rates sortable by a unique neighbourhood identifies

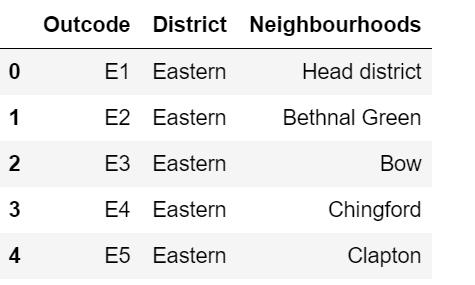
**Data**

Building from the requirements identified above data sets have been systematically selected and wrangled to ensure they can each provide added information to answer the projects goal. All tables bellow has already been cleaned with feature reduction for ease of viewing of the relevant information. Additional features have also already been added but noted under their respective sections.

Requirement 1: A way to identify neighbourhoods at a granular level

Source: Royal Mail

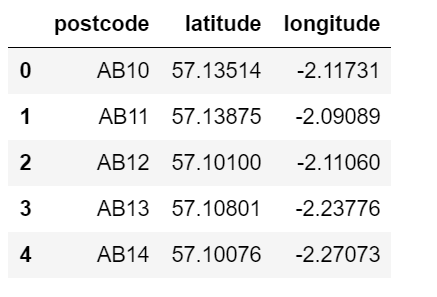
After some data manipulation and exploratory review, I determined that using the UK standard Out Code which is the prefix two – four digest of the post code which are unique to a given geographic location. During my discovery I determined that out code was the best medium as postcode was to granular and was difficult to matching other data sets. And borough level was not granular enough to provide detail. Most data set I selected were either arranged by neighbourhood name (which have a unique pre out code) or were sorted by out code.



Requirement 2: A way to geolocate these neighbourhoods

Source: NOMIS

In order for the venue data and graphing of the neighbourhoods I needed a way to match the Out code to latitude and longitude fortunately I was able to find one such repository provided by the UK royal mail group providing the Out code and the latitude and longitudinal centres for each.

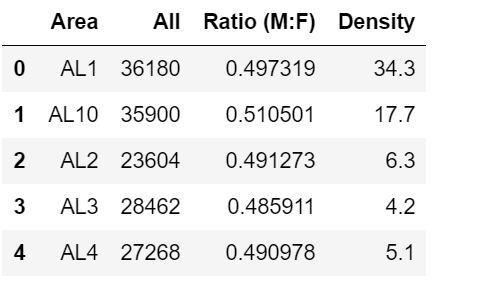


Requirement 3: Neighbourhood demographic data to help with clustering analysis

Source: UK Census

To further develop the clustering model, I wanted to incorporate demographic information to develop more detailed clustering features and allow for further data metrics if I decide to implement a dashboard visualisation for this project.

As neighbourhoods are not similar in means of density, population, gender split incorporating these metrics will allow me to standardized absolute values or venue counts with a metric of venues / person or venues / density. Added feature ratio (M:F) previous table gave total number of males, females and both together, I removed the gender specific values in exchange for ratios with can be multipled out to find exact gender value as needed. Density is in relation to the 1000 people per hectare of land.



Requirement 4: Data base of venues that can be grouped by geolocation

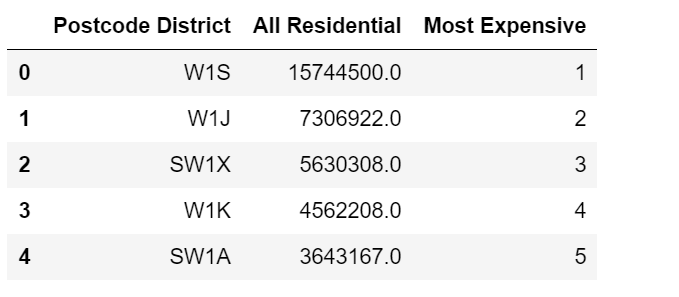
Source: Foresquare

Foursquare offered an ideal APi access to venue data and continuing updateable information if this programme is going to be used for more than just my own needs. Foursqure matches both my needs of providing geolocation for specific venues and the ability to perform radius searches from neighbourhood central coordinate information iteratively.

Requirement 5: Data repository of housing prices sortable by a unique neighbourhood identifies

Source: UKCrimeStats.com

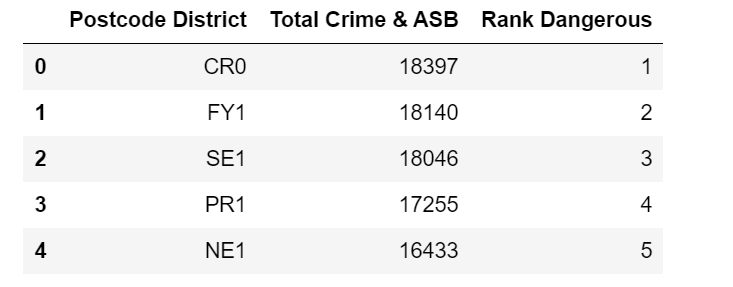
To be able to fine the second requirement for my search I wanted to look at average (mean) selling price for residential prices in 2018 in each postcode across the UK and assign them a ordinal ranking of most to least expensive postcode. Most expensive is a ranked ordinal value assigned from absolute mean value of residential sales 1 – 2200+.



Requirement 6: Data repository of crime sortable by a unique neighbourhood identifies

Source: UKCrimeStats.com

To be able to fine the final requirement for my search I wanted to look at average (mean) selling price for total number of crimes in 2018 in each postcode across the UK and assign them a ordinal ranking of most to dangerous postcode. Most Dangerous is a ranked ordinal value assigned from absolute total number of crimes in a area 1 – 2200+.



**Methodology**

**Result**

**Conclusion**