**Clustering London Venues and Analysing Correlation to Property Prices**

**Introduction**

London is one of the largest metropolises in the world. With a population of over 8 million people living in a space of 1572km2, this bustling city is full of venues ranging from restaurants to night clubs. As a result, each neighbourhood has grown a distinct character which often drives the types of venues that dominate the area. For instance, we would expect to find more night clubs and bars in vibrant Eastern locations, such as Shoreditch, when compared to the family friendly areas of the West, like Hammersmith and Fulham.

Venues often reflect the lifestyle of local residents and can dictate the status of tenants that Landlords might be exposed to. This is crucial for property owners looking to live in areas which complement their personality or landlords targeting specific demographics (e.g. professionals in Canary Wharf). In turn, this drives the business decisions of commercial properties who are trying to provide services in line with the needs of the local neighbourhood. Therefore, the distribution of venue types across different regions of London is particularly important for citizens looking to live or invest in London.

As a lifetime resident, analysing the distribution of venues in London’s neighbourhoods is particularly interesting since it provides insight into the distinct character of each area. The project also intends to explore potential correlations between venue types and property prices. This information can be used to establish which venue distributions are associated with higher property prices as well as providing a basis for more accurate property price predictions.

**Data**

In order to complete the analysis, three main steps and datasets will be required.

First, London needs to be split into relevant regions in order to compare the distribution of venues and their correlation to property prices. The city is already conveniently sectioned into postcodes with clearly defined boundaries and addresses. A map and csv list of postcodes, including their comprised neighbourhoods, can be found at: <https://www.doogal.co.uk/london_postcodes.php>. The Data will be turned into a pandas Dataframe and then cleaned to display boroughs using the first 3 characters of each postcode, the longitude and latitude values of their centres, and the neighbourhoods contained within each borough.

Second, a list of venues in range of each borough will be required for clustering. This can be done using the **Foursquare API** and defining an appropriate radius within which to query. Once a list of venues has been obtained, exploratory analysis can be performed to see what kind of venues are most abundant in different regions.

Third, property prices will be needed in order to explore potential correlations. Data on average property prices divided by postcodes is publicly published by the UK government and readily available from many sources, for example from Rightmove at the following link: <https://www.rightmove.co.uk/house-prices.html>. This information can be combined with the previously clustered groups to perform an analysis and see whether (i) the property prices of the postcodes grouped by cluster are comparable to each other and (ii) whether there is a correlation between the types of venues and property prices.