Introduction/Business Problem

The experience of moving home is often a stressful one. Whether you're relocating to a new city or just a new neighborhood, there's plenty to manage and organise. Arguably the most draining activity of all is the search for your new dream home. I've experienced this myself more than once, most recently when I've decided to relocate from the small, peaceful town of Basel, Switzerland to busy, bustling, gigantic London. Where does one start to look for a home in a place like this?

There are the obvious constraints of the workplace commute time and price range, but that still leaves more options than one can ever hope to find the time to go through. Despite having a fairly clear idea of what I was looking for, not only in my ideal home but also in the area I wanted to live in, I could only check how well each property matched by going over them one by one. The number of choices was overwhelming and exhausting, all the while I couldn't shake the feeling that I was missing out on the perfect property.

What if we could make this process easier by narrowing down our search and only focusing on a subset of neighborhoods that fulfill criteria of our choice beyond location?

My project will use London as a proof of concept for how this could work: use venue information to cluster similar neighborhoods, then apply additional filters, such as distance from a given location and other random desired features (eg. green spaces, restaurants, ethnic shops), in order to narrow down the selection to only the few most suitable areas that fulfill these criteria. This tool can benefit anyone looking to move house, making the process of finding a new home easier and more enjoyable.

Data

To identify the required data, some preparatory research is needed. We first need to identify the optimal level of territorial subdivision within the city of London. A quick search reveals that the borough compartmentalisation corresponding to London postcodes is at just the right level of granularity for our scope. The list of boroughs is available for scraping on Wikipedia. Next we will need to retrieve the coordinates of each district using a geolocation tool before we can employ the Foursquare API to complete our dataset with venue information for each district. This will provide us with all the necessary information in order to cluster similar neighborhoods based on their venues, then apply additional filters to identify the most suitable areas according to our criteria.

On a side note, London is a dangerous city. Information on crime rates in each borough is available on a number of websites. We can use this information to assign a safety score to each neighborhood and include this feature in either the clustering criteria or post-clustering, when filtering for most suitable areas to live in.