Capstone Project: Prime coffee shop location in London

1. Introduction

The Problem: ***Where should I open a new coffee shop in London?***

If I were interested in opening a new coffee shop in London - a bustling city with plenty of independent cafes and restaurants scattered over many neighbourhoods - where would the best location be? In this project, I'll answer this question using location data from the Foursquare API.

In order to find the perfect spot, we need to consider criteria that might be markers of a "good" location. One clear marker could be locations that have few existing coffee shops, which would minimise competition. However, locations might not have coffee shops for good reasons (lack of demand/passing trade, etc.). So we should probably consider some other factors too. Another marker of a good location could be proximity to universities — students like caffeine! Finally, we would want a good amount of passing trade, so proximity to retail and other attractions, such as museums and theaters, should also be considered. All in all, the following criteria should help prospective coffee shop owners to choose a good location:

1. Few existing coffee shops (low competition)
2. Near students (caffeine demand)
3. Near retail, museums or other attractions (passing trade)

In this project, I will highlight the best possible neighbourhoods by exploring and clustering locations within the central London area, using the above criteria to select a group of candidate locations.

2. Data

The data used for this project comes from Wikipedia (London locations) and the Foursquare API (information on surrounding venues). Together, the data include London postcodes, with their latitude and longitude, as well as information about the number of coffee shops, universities, shops, and attractions for each location collected using Foursquare.

- London postcodes are scraped from this Wikipedia page: https://en.wikipedia.org/wiki/List_of_areas_of_London. The data will be filtered to find all of the neighbourhoods codes within the "London" postal town.
- I'll use the Foursquare API to run a series of searches for each neighbourhood, collecting the number of venues within different search categories (coffee, universities, shops, entertainment) per location.

The Foursquare data will be used to cluster my London locations according to their venue distribution (i.e. the frequency of coffee shops, universities, retail, and entertainment venues that exist in each area).