Finding a home at London

IBM Professional Data Science Specialization Capstone

Alexandros Spiliotis

Description/Background

As the trend of moving towards large cities, where the well-paying jobs are, continues, finding the right place to live at a city gets more and more difficult.

The goal of this work is to create a useful tool for people who are considering moving to the Greater London Area and want to choose a suitable neighborhood for them to move at. Of course, the notion of "suitable" depends one each house seeker's demands. Some crave lively neighborhoods, full of places to eat, drink coffee and have all kinds of fun right outside their doorstep; others prefer quiet neighborhoods, and instead commute to neighborhoods which can offer amenities that theirs don't have; and others just want a place close to their job, or just an affordable place. Having taken this into consideration, this work has been designed in a way such that the results are matched to the demands of each user.

To achieve this, a cluster model is created that categorizes London neighborhoods on the basis of certain real-world parameters. The number of clusters, as well as the weights for each parameter, can be modified by the user.

The parameters used are the following:

- Restaurants: In this category, all venues that are generally categorized as Restaurants, such as Italian or Japanese Restaurants, but also steakhouses and gastropubs.

- Food stalls: This category includes food places other than restaurants, such as burger joints, food stalls etc.

- Nightlife: Bars and Nightclubs.

- Stores: Supermarkets, convenient stores, but also department stores.

- Cafes: Coffee and tea houses and other kind of cafes.

- Art Venues: Art galleries, theaters and museums.

- Gyms: Gyms, spas, pools etc.

- Outdoors: Parks and game courts.

- Distance From Target: The distance from a specific location that the user will determine, be it their job, their University, the house of a relative etc.

- Housing Values: The median price of a house at each neighborhood.

Data

Housing prices per neighborhood

For this purpose, a dataset from data.london.gov.uk was used. The dataset contains median and mean house prices per neighborhood of the Greater London area, from December 1995 to December 2017, and the corresponding number of house sales. Data were recorded every trimester. As mentioned in the website, "...property prices (were) calculated by the GLA from Price Paid Data published on Land Registry website".

The dataset can be found at https://data.london.gov.uk/dataset/average-house-prices.

London venues

To gather the venues of London, data from Foursquare API were downloaded, by calling all venues at a certain radius (modifiable by the user; default is 500 meters) from the center of each neighborhood.

FOURSQUARE