

Searching for A Living Place for a Student New in Boston Using Machine Learning Techniques

Usama Bin Shakeel
Coursera

Idea:

Boston is a city of universities. Students from all over the world come here to study. Moreover, due to being home for a number of tech universities, students from all over America come here for universities. These students face a problem in the form of finding a suitable neighborhood for residing, which has all the living essentials in walking distance. So, in this project, using machine learning techniques we found the suitable neighborhood for students to live in.

Dataset:

We used the dataset of neighborhoods obtained from <https://data.boston.gov/dataset/boston-neighborhoods>. Then we searched for the corresponding longitudes and latitudes of the neighborhoods and manually entered them in the data frame by introducing new columns. This was a tiresome process but was necessary as we weren't able to find any dataset containing longitudes and latitudes of neighborhoods of Boston.

Methodology:

First of all, we deleted all the unnecessary columns from the data frame which were of no use to us. Then we found the longitude and latitude of Boston using certain python libraries. Then we popped it up on the world map in a zoomed form. The neighborhoods were visible in the form of dots on the map. A thing to remember here is that we were restricted to use foursquare API in order to complete our project.

Next thing we did was to make a foursquare API call using our given credentials. This gave us venues for each neighborhood and their corresponding categories. Then we searched for total number of unique categories and after doing one hot vectorization for common venues we searched for fifteen most common venues for each neighborhood and put them in the form of a data frame.

Now came the moment of applying our machine learning skills. The machine learning algorithm that we used was K-Means clustering, a clustering technique commonly known among machine learning enthusiasts and which is very efficient to use. The number of clusters was kept to five which was an optimum number and the clusters were visualized on the map.

Results:

The cluster 1 resulted in the largest number of neighborhoods with twenty neighborhoods in it. The most common venues in the neighborhoods of this cluster seem to be restaurants.

Cluster 2 contained only one neighborhood with store and museum as the most common venues.

Cluster 3 also contained only one neighborhood. The most common venues seem to include small food shops for take away.

Cluster 4 also contained only one neighborhood which seemed like a real neighborhood with residential facilities.

Cluster 5 also consisted of only one neighborhood with small food shops and other facilities as well.

Discussion:

If we closely observe the neighborhoods in all the five clusters, we can see that the most suitable neighborhood for a student to reside in is West Roxbury. It lies in cluster 4. This neighborhood has all the necessary facilities which a residential neighborhood should have in walking distance such as bank, gym, train station, pharmacy etc. All the other neighborhoods in other clusters seem more to have a saturation of restaurants. So, it is recommended based on the results that a student should search for a living place in West Roxbury.

Conclusion:

Hence using k means algorithm and making use of foursquare API, we were able to devise a result. Though the result could have been quite improved using different features if we were not restricted to use foursquare API, still we obtained significant results which can be trusted.