

IBM Applied Data Science Capstone

Opening new Indian restaurant in Atlanta: An exploration of neighborhoods in Atlanta, Georgia, USA

By,

Kanchan Basnet

June 2019



Data source

The Wikipedia page https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Atlanta was used to find and extract the name of the neighborhoods in Atlanta, Georgia. Once the neighborhoods' list was extracted by using web scraping technique, geographical coordinates for those neighborhoods are required to plot them in a folium map and also to get the name list of the venues in the particular region.

Data extraction methodology

A list of 123 neighborhoods was extracted from Wikipedia page using web-scraping packages BeautifulSoup4 and python-requests. The list was then converted to pandas data frame and cleaned so that it was ready for the geographical coordinates to append. Python geopy.geocoder package was used to search the longitude and latitude of the neighborhoods in Atlanta. The obtained geographical data was then merged with the names of the neighborhoods in the pandas data frame.

Foursquare API was used to extract the venue data of a particular neighborhood in Atlanta. According to Foursquare, more than 150,000 partners and developers are using the webpage. Once the venues data was obtained, data cleaning and data wrangling was done, which makes the data ready for visualization and clustering. Folium package

was used for the map visualization. Towards the end, machine learning technique k-means clustering was used to cluster the extracted list of Indian restaurants into four different clusters. Each cluster was then studied, and the decision for the appropriate spot where chances of business success would be maximized was recommended.