

Prime locations for a new restaurant chain in Washington D.C.

Data

The datasets of this project include the information of neighborhoods, local venues (e.g., coffee shops, grocery stores, fast foods), hotels (e.g., rating and reviews), and safety (e.g., crime reports). If necessary, the annual household income of each neighborhood will be included in the analyses. The information of local venues and hotel chains will be attained by using Foursquare API and the crime reports of DC area is publicly available for download.

These datasets will be explored individually first in order to clean up the dataset before any analysis. Then, if necessary, some datasets (e.g., local venues and safety datasets) will merged in order to further explore the segregation among the neighborhoods. Based on the restaurant types, we can use different python and R packages (e.g., Numpy, Matplotlib, MASS) to cluster specific neighborhoods into different categories and find the specific locations for the new restaurant.