Introduction:

It is always desirable to open a business at a place where the supply is low and the demand is high but it is hard to figure out where it is. Many things have to be taken into consideration, such as the demographic and economic characteristics of the area, competition in the area and other business that might complement your own. One can’t easily go to every corner of the city to collect all the information. Fortunately, with the help from the city’s geographical data and the venues data from Foursquare, we are able to solve the problem without going on the field. This project will use the city of Cleveland as an example to demonstrate how data science can help people who are looking to open a common business like a pizza place or a sandwich place find the ideal position where the demand is high and the supply is low.

This process can be broken down to two major steps. First, one can acquire data from Foursquare, find out the popular business in all neighborhoods and group the neighborhoods into clusters where the neighborhoods in the same cluster have similar venues. It provides a pattern of venues that are widely successful in many neighborhoods. Then we are able to compare the venues between neighborhoods inside the same cluster and find out whether a common business is missing in a neighborhood. If a common business is missing it means the supply is low and we would know about the demand by checking the other venues in the neighborhood.