Introduction – Business Problem – Data Set

By Tom Roush

I will use two data sets to solve this problem. The first is a data set provided by the New York City Department of Planning that uses centroids to report on population density. This data set uses densities to report on neighborhoods, so it's perfect for what I want to do. I can limit the reported neighborhoods to the appropriate borough, and then it will place a pin to denote where each neighborhood is located. HEB is interested I Manhattan, and so the data will show the neighborhoods of Manhattan as denoted by the center of population density. Perfect!

The next data set is Foursquare, which I can link to the Department of City Planning data via it's latitude and longitude. Then, I will have a great working map of the neighborhoods, as reported by the first factor HEB cares about (population density) overlapped with the grocery competition.

With these two data sets, we can find the weak point in the market and selection the neighborhood where HEB can launch a flagship store in Manhattan.