Report for IBM Data Science Capstone
Richard Anastasi

I am a software developer looking to move somewhere in the Northeast of the United States. My main two cities of interest are Boston and New York City. One major consideration for where I want to move is the density of museums in either city. Whichever city has the best museum district is the one I am more likely to move to.

The data used to resolve this issue came from several sources. The first listed website gave me a list of all neighborhoods in Boston and their corresponding zip codes. The second gave similar data for New York City.

http://archive.boston.com/news/local/articles/2007/04/15/sixfigurezipcodes_city/

https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm

After reading in the New York and Boston data into their own dataframes I first had to clean the data. Boston zip codes have a leading zero which disappears when they are read as integers, so they needed to be cast to strings and corrected. Geolocator was used to obtain the geolocational coordinates of New York City and Boston to be used later in the project. A csv file was then imported with every U.S. zip code and its corresponding geolocational coordinates. This csv was then merged with the NYC and Boston dataframes with zip code as their index (after some more cleaning). The foursquare api was then used to discover what proportion of venues in any given NYC and Boston neighborhood were art museums, history museums, and untagged museums. The data was grouped by neighborhood and we found the data we were looking for.

In NYC Bronx Park and Fordham have the highest proportion of museums, but the density of NYC made it so the density of museums in the neighborhood of East Fens in Boston was much higher. With a proportion of roughly 0.156 museums per venue, it is the clear choice for us!

We can conclusively determine that given our criteria that East Fens is the neighborhood we want to move to!