

Coursera Data Science Capstone Project Report

Business Site Selection with Foursquare and Python

Problem Definition and Analytic Approach

Dec 19, 2018
Bryant Sheehy Jr.

Introduction

For my capstone project, I chose to use the Foursquare data set and Python data wrangling and mapping tools to help solve the problem of where to locate a new business in the city of Chicago. I live in the Chicagoland area, so I'm familiar with many of the neighborhoods of the City and the general layout. I also happen to have some experience in a previous life as a commercial real estate broker, so I know a bit about site selection.

When deciding where to open a new retail or service business, there are a number of considerations to think about. First of course, you want to think about who your potential new customers are and where they might be located. Then you want to look at similar businesses already located there to get a sense of what the potential competition might look like.

To keep things simple for the purpose of this project, we're going to address the first consideration by focusing the study within the city limits of Chicago. Chicago is a very diverse and relatively dense urban area. It is made up of seventy-seven distinct communities or neighborhoods with a wide variety of cultures and income, similar to boroughs in New York. So there should be potential customers of all types for just about any kind of retail or service business here.

The Problem

The problem we will try to solve is where to locate a new business such that the competition will be minimized, all else being equal.

We will try to solve this problem by identifying communities within the City where there are fewer potential competing businesses and therefore potentially unmet needs. Within the Jupyter notebook, a user looking for the right location to open their business would be able to enter one or more business categories to search, and see a list of how many businesses in this category are located in each community along with a couple of maps showing which communities have higher or lower concentrations of these types of businesses based on locations per capita and/or locations per square kilometers.

The Analytic Approach

The Foursquare Search API provides a convenient method of searching any urban geographic area and pulling up a list of most businesses located in that area with each business assigned to one or more very granular business type categories. Bing provides an API that allows you to pull the geographic coordinates for any location in the world that can be described in terms of a place name. Wikipedia provides a good description of each community within the City of Chicago which includes the number of people living there and the size of the geographic area in square kilometers. So we will use data from these three sources.

From Wikipedia, we will pull the names of each Chicago community along with its population, geographic area and income statistics.

From Bing, we will pull the geographic coordinates for each community.

From Foursquare, we will pull the names, categories, addresses and geographic coordinates for each business located in the Chicago city limits. We will then allow the user to enter in one or more Foursquare business categories to search for and display a couple of maps, choropleth and marker clusters, that indicate which communities have lower densities of the chosen business type, higher or lower median household income and where the potential competition is located.

The idea is to help the user find areas in the City where customer income is adequate, competition might be lower, and where they might be able to find underserved geographic holes in the market.