IBM Applied Data Science Capstone Project Proposal

Brian Jairam

February 7, 2020

Choosing location for new restaurants is tough. I propose a method of selecting the most advantageous U.S. cities to expand to by clustering cities with respect to the number of restaurants of similar type, then selecting the cities with the least of such restaurants per population.

1 The Problem

Suppose we are a the owner of a thriving U.S based restaurant chain with ambitions of expanding beyond the confines of our midsize city. Motivated by passion, fame, and of course, cold hard cash, we set out to raise the capital needed to open a brand new location. There is but one problem: where should we open our restaurant?

Our first instinct is to go somewhere big. More people means more exposure, right? However, our cousin (who works in the kitchen) reminds us that cities like New York or Los Angeles, while packed with potential customers, are also packed with stiff competition. However, he agrees that my mother's suggestion of Nampa, Idaho is not ideal either. Good, we do not really want to go home any time soon anyways.

So armed with a little knowledge of data analytics we picked up on Coursera, we decide to choose our target city a little more methodically...

2 The Solution

We will use a dataset of the 1000 largest US cities hosted on OpenDataSoft. Foursquare will be used to generate the types of 1000 restaurants closest to the city core. Using this information, the cities will be clustered using a k-means algorithm. Now, we will inspect and label these clusters, choose the clusters in which restaurants of our type are most prevalent. We will then look for markets in those clusters that have room for our franchise to blossom (high population, low competition).

We will then let our cousin pick from our shortlist... in lieu of giving him a raise. We need to save some money right now after all!