**Introduction**

This is the report for the IBM Data Science Professional Certificate capstone project. Thus far we have completed eight courses covering the fundamentals of data science, a range of core algorithms (such as K means, support vector machines, linear and logistic regression), and to how to use Python and SQL to visualize and solve specific problems. In this assignment we have been tasked with using Foursquare data to solve a real-word problem of our choosing.

The problem I have chosen is identification of potential locations the would be ideal for a new coffeeshop based on frequency of neighboring establishments.

**Problem Statement**

We have all seen countless failed restaurants, coffee shops, and the like in our travels. Most can probably name a few that are likely to fall away in the coming years. While there are many factors that go into a successful business, and restaurants are no different, without an ideal location a restaurant will not be successful. If there are too many well-established competitors a new coffeeshop will fail to draw business and will not stand a chance. In this project we will identify areas around Atlanta that are deficient in coffee (and/or other establishments if need be) to find low competition areas a coffeeshop might thrive in.

The target audience is those who wish to identify a good location to start a coffeeshop business.

**Data Collection**

As mandated by the guidelines given for the assignment, Foursquare will be the primary source of data. It is straightforward to pull zip code-associated venue category data from foursquare, as shown in the following image.



Neighborhoods will be grouped using K-Means, establishment types will be grouped to improve the usability of the data, and correlations will be found between coffeeshop frequency and the frequency of other establishments so as to identify areas with a clear coffee deficiency relative to other well-correlated establishments.