**Optimized Rating Model Part I**

**An Exploratory Analysis**

**Introduction:**

Before grabbing a bite to eat, restaurant patrons often reference user ratings to make decisions on where they go. The problem is these raw ratings can at times be incomplete or biased, which is a disservice to the restaurant owner and the patron. Foursquare and other clients that publish user ratings for restaurants have interest in building an optimized rating model to remedy this problem. As a first step, we would like to have available an exploratory analysis on which variables would be most useful to include in this model.

The end-result of an optimized rating model is likely a classification model that scales from 0 to 100% (5/5 stars) that closes in on the actual user rating as the number of reviews approaches a certain threshold. The scope of this project is to define which independent variables to consider in this model. A/B testing for all potential variables with an incomplete dependent variable could be tedious and incomplete. Alternatively, we can use K-means clustering to give us more insight into which variables to consider in the final model.

**Data:**

We will focus our research on restaurants listed on **Foursquare** ([developer.foursquare.com/](https://developer.foursquare.com/)) in the Chicago major metropolitan area. Along with user ratings, number of ratings, cuisine type, location, and other variables we can obtain from Foursquare, we will reference demographic information about the tracts in which the restaurants are located from the **Federal Financial Institutions Examination Council** (FFIEC - [www.ffiec.gov/](http://www.ffiec.gov/)). A census tract *is a geographic region*defined*for the purpose of taking a*census (<https://en.wikipedia.org/wiki/Census_tract>). FFIEC search results are based on 2015 Census data.

For example, one data point is **Chapati Mediterranean Restaurant** with a user rating of **6.5**/10 given **8** total ratings and **Moderate** affordability according to Foursquare. The tract this restaurant is located in has a population of **4,888**, an **Upper** income level (definitions for income levels here: <https://www.ffiec.gov/census/htm/2015CensusInfoSheet.htm>), and a median household age of **30** according to FFIEC.