**Data**

The correlation between weather and the restaurant business has been widely observed. A study published by Ohio State University (Milos Buijsic et.al) suggests that the weather not only affect’s a restaurant’s volume of customers, but also how much they enjoy their experience at a restaurant. Moreover, according to the National Restaurant Association, “More than 90 percent of restaurant operators indicate that changes in local weather conditions affect their sales and customer counts.” We will be using a 5-day forecast of the temperature and humidity for our project. There will be a temperature and humidity observation recorded every 3 hours so in order to obtain the total average 5-day temperature/humidity, we will need to average 40 data points for each category.

Apart from weather data, which we will be drawing from OpenWeatherMap’s API, we will be merging this dataset onto Foursquare’s location dataset. We will use Foursquare to obtain a list of all venues in various neighborhoods to determine which neighborhood has the highest restaurant frequency as well as total restaurants. Industrial organization, a field of economics that studies the theory of the firm, suggests that in a given geographical plane, similar firms will eventually start to clump together in at a pareto-optimal focal point in order to optimize revenue. This is why you will see stores that sell similar products located in very close proximity to each other. With both weather, location, and venue data, we will help restaurant owners decide whether or not they should open their store in New York or Toronto.

As a note, the most extensive and cost-free API forecast we can obtain from OpenWeatherMap is a 5-day forecast which in theory shouldn’t be sufficient enough to base a business decision off of. Although a yearly aggregate of historical weather data is preferred, we are not able to obtain it. Also, there are many other factors that business owners need to consider when opening a restaurant such as per-capita income, local supplier costs, and the frequency of the number of restaurants with similar cuisines. It is also important to recognize that neighborhoods within the same city should have very similar temperatures and humidity since they are geographically very close, so it is hard to exploit weather data here. Thus, we can only use weather data to analyze differences between cities and not inter-city neighborhoods. A better use of this project would be to analyze regions of a certain country, but this is something that is beyond the scope of this course.