IBM Data Science Final Report – Week 4
Faddy Kamel

1. **Introduction: where you discuss the business problem and who would be interested in this project.**

   In this project I will attempt to find a relationship between the most common types of restaurants (including coffee shops, bars, etc) in a city and the city's obesity rate. If the impact on the obesity rate of a new restaurant opening can be predicted, this problem can be tackled more effectively. Since it is known in the medical community that nutrition is at least 50% responsible for the overall health of people, finding a correlation here can help deploy health improvement efforts more effectively.

   The main stakeholders of this study would be doctors, health officials, fitness companies, and lawmakers. Doctors & health officials will have more information to predict the trend in the population they are treating and therefore can be more accurately proactive in preventing obesity. Lawmakers can use that information perhaps to pass legislation limiting the density of unhealthy restaurants or creating community initiatives to promote healthier lifestyles. Finally, fitness companies can leverage a population's obesity rate and willingness to reduce it to generate revenue.

2. **Data: where you describe the data that will be used to solve the problem and the source of the data.**

   The obesity data will come from governing.com, which itself took it from the CDC. Unfortunately, the original CDC database does not seem to be accessible. The dataset contains a list of 192 American metropolitan areas and the following data for each:

   1. Normal weight (% of population)
   2. Overweight (% of population)
   3. Obese (% of population)
   4. No physical activity (% of population)

   The cities dataset will need to be cleaned before using it for API calls to make sure the final dataframe can be readily analyzed & used to visualize the results.

   The restaurant data will come from the Foursquare API. I do not need coordinates for each city since I can use the "near" parameter when executing calls. For each city, I will retrieve:

   - Venue name
   - Venue location
   - Venue category

   I will use the results of the API calls to build a dataframe of the count of each of the most common restaurant types in each city.