Katelyn Veyna
Springboard
Data Science Intensive
September 2016
Capstone Project Proposal

1. What is the problem you want to solve?

   I would like to be able to predict restaurant inspection scores of new restaurants or existing restaurants based off of Yelp reviews for the city of Las Vegas in Nevada and already existing restaurant inspection scores.

2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

   My client could be the restaurant. It would care to solve this problem so that it could know in advance where the cleanliness of the restaurant lies according to its customers. Based on the predicted scores, the restaurant could decide if they need to improve its procedures for cleanliness and health safety of the restaurant or maintain what it has going.

   Another client could be Yelp itself. Yelp could extend this problem and solution to other cities that provide restaurant inspection scores. The company could provide this information on their website for restaurants that don't yet have an inspection score or just offer this predictive score in addition to existing inspection scores. The predictive score could even be better than an existing score since it was acquired using other user reviews and other restaurants in the city.

3. What data are you going to use for this? How will you acquire this data?

   The data will be restaurant inspection scores for restaurants in Las Vegas and Yelp reviews for these restaurants in Las Vegas.

   Las Vegas has an open data source for government datasets, so from here, a file for restaurant inspection scores can be accessed.
   Here is the link: https://opendata.lasvegasnevada.gov/Public-Safety/Restaurant-Inspections/q8ye-5kwk

   Yelp has publically available data for several cities from their Yelp Dataset Challenge. Las Vegas is one of the cities they have extensive data on.
   Here is the link: https://www.yelp.com/dataset_challenge/

4. In brief, outline your approach to solving this problem (knowing that this might change late).
   First, I will need to limit the Yelp data to only restaurants in Las Vegas. Next, I will play with the data to see if I can see any visual or statistical patterns. From there, I might see if certain keywords in a review could indicate restaurant inspection scores. Eventually, I will want to implement machine learning using either a supervised or unsupervised approach or both to get the predictive aspect of the problem.

5. What are your deliverables? Typically, this would include code, along with a paper and/or slide deck.
   A deliverable is going to be a paper describing the problem, my approach to solving this problem, the results I obtained, and then other possible problems or approaches that can be used to extend this project. Also, there will be a slide deck outlining the paper that can be used to give a presentation on my capstone.