Capstone Project Proposal

Student: Christopher Jose
Mentor: Raj Bandyopadhyay

Goal:
Determine the main factors that affect Yelp scores of Starbucks shops.

The client and why they should care:
My client is Starbucks. By determining the factors that lead to a higher Yelp score, Starbucks can:

    a. Focus on improving these factors to improve the guest experience for existing shops.
    b. Keep these factors in mind when building new shops so as to best position them for success.

Where the data is coming from:
I am using the freely available dataset that Yelp provides as part of their "Yelp Dataset Challenge".

My approach to achieving the goal:
I will import the Yelp dataset into pandas and convert it into a DataFrame. Then I will build a new DataFrame containing only Starbucks locations, which I will use for analysis.

Dependent variable data will consist of the Yelp review score for each Starbucks contained in the Yelp dataset. Independent variable data will potentially consist of variables like location, review text content, etc.

I will carry out exploratory data analysis to determine potential independent variables, such as by creating visualizations like box plots to see if there is a relationship between the considered independent variable and Starbucks review scores.

I am going to generate multiple prediction models to predict Yelp scores and will select the model that best predicts Yelp scores. Within this model, I will select statistically significant factors that have the largest effect on Yelp score. I will then recommend these factors as deserving consideration in Starbucks' store improvement efforts.

Deliverables:
The report and the code.