**Machine Learning Mock-up:**

Dataset: Yelp Academic Dataset

Analysis: Machine Learning

Type: Supervised

Branch: Classification model

Hypothesis: Non-American traditional restaurants receive a higher star rating in the Western States. (edited)

Task: Use supervised machine learning to predict whether a new non-American traditional restaurant will receive a higher rating or low rating within the Western States.

Using the locational data included in the dataset will help with the visualizations of low vs high rated non-American traditional restaurants on the map.

What constitutes a high rating versus a low rating will need to be defined.

Models to be explored:

# Logistic Regression model

# Support Vector Model

# Random Forest Model

In the event we run into the problem of class imbalances, then the following models can be explored to arrive at the best model that yields the best results:

# Random Over Sampler and SMOTE algorithms for over sample

#Cluster Centroids algorithms for under sample

#SMOTEEN for both over and under sample  
#Balanced Random Forest Classifier and Easy Ensemble Classifier for bias reduction