Ever since I starting coding, Machine Learning has intrigued me. It is an increasingly relevant and growing specialization for a software developer, but I never knew where to start. Luckily, I was offered a tremendous opportunity to take part in a challenge held by CUNY and sponsored by Point72 who gave us the dataset we’d be working with and basic code to get started. The data was based on inspections held on the restaurants of New York City and any violations found.

The goal was to use a Machine Learning Model to predict whether a given restaurant would pass or fail based off previous data we taught the Model with. Then we would check how confident the Model was in its predicting and penalize it accordingly, and so our final calculated score was based off of the Log Loss.

The competition was held on Kaggle and the top 5 teams with the lowest scores and best presentations would be chosen to present their powerpoints to their fellow students as well as professionals in the data science industry. I was one of those chosen and was able to give a presentation on my findings and how I was able to manipulate and feature engineer the data given in order to lower the Log Loss.

Here are the slides I presented, and the brief overview of the project