**Neil Oza**

(770) 710-8913│neil .s.oza@gmail.com

GitHub: [Link](https://github.com/neiloza/Springboard) │LinkedIn: [Link](https://www.linkedin.com/in/neil-oza-9902438a/)

**Identifying High Risk Loans**

* Obtained current Lending Club loan data from Kaggle. Opened the data in Jupyter Notebook as a Pandas dataframe and looked over the features for preliminary analysis
* Identified a question to answer with the data: what loans that are currently current are at high risk of default? Determined the column ‘loan\_status’ suitable as a target variable for this question
* Performed EDA to understand the distributions of features. Identified columns with large numbers of null values, created histograms to model the distribution of the different features, and identified features highly correlated with the target variable
* Filtered problematic columns out of the dataset to create a feature set. Removed any column with a large percentage of null values, did not lend themselves to encoding (such as zip code), or were too highly correlated with the target variable (such as recoveries). Created dummies to encode the categorical variables, and created two versions of the featureset that used fillna(data.mean) and fillna(data.median) to handle null values
* Sampled 100,000 records from the data to train and evaluate the models. Split the sampled data into training and test sets
* Trained SVM, Naive Bayes, Random Forest, Logistic Regression, and Gradient Boosting classifiers on the data and constructed performance metrics for the various classifiers, including ROC curves and precision-recall curves. Normalized and standardized feature sets as appropriate
* Used cross validation and Hyperopt to tune hyperparameters. Constructed a lift curve to highlight the gains from the models

**Denver Crime Data Analysis**

* Gathered crime data from Denver’s city website. Opened the data in a Jupyter Notebook using pandas, and began preliminary analysis
* Reached out to Denver officials to clarify several features. Determined the dataset contained four features represented in multiple forms. The features are type of crime, location of crime, time of crime, and Offense ID
* Used data cleaning techniques to check for inconsistencies within the data. Determined that several timestamps contained errors and sent a report to Denver officials
* Created visualizations to identify crime trends in the city. Organized crimes by type, district, and time. Determined that auto-thefts have increased in prevalance over the past five years, that drug and alcohol crimes are concentrated in one particular district, and that many types of crimes demonstrate seasonal behaviors. Conducted hypothesis tests to verify the results
* Used the python library Folium to create a heatmap generating function to visually overlay crime locations on a map of the Denver. Added filtering capabilities to allow Denver officials to focus on crimes by type, time of day, and location
* Created a report summarizing the findings of the analys

**Education**

**Springboard Jan 2019-Present**

Data ScienceProgram

-Online program consisting of 600+ hours of curriculum, with 1:1 industry expert mentor oversight and completion of 2 in depth capstone projects and several mini projects

**University of Georgia Aug 2012-Dec 2015**

B.S. Mathematics

GPA: 3.34

**Work History**

**Private Mathematics Tutor Mar 2017-Present**

- Help students learning math and science subjects such as Calculus, Probability, and Chemistry understand the material

**Colonial Life Agency Sales Representative Mar 2016-May 2017**

-Worked with local business owners to provide comprehensive insurance packages for their employees