* Description of preliminary data preprocessing
  + In our preprocessing, the only thing we changed was converting ‘Arrest’ and ‘Domestic’ columns into ones and zeros and not encoding the other ‘object’ type columns. In our csv file, we also adjusted the ‘Safety’ ratings by increasing range of ‘safe’ including misdemeanors (rankings of 3 or below).
* Description of preliminary feature engineering and preliminary feature selection, including their decision-making process
  + Because we introduced our own ranking system, we determined that encoding our dataset was not necessary. We also used StandardScaler to normalize the data. For our feature selection, we filtered out more unnecessary columns. Specifically, we removed, ‘Population’, ‘Latitude’, ‘Longitude’, and ‘Ward’.
* Description of how data was split into training and testing sets
  + In our original model, we split the dataset using the default settings of 75% training and 25% testing. In our neural network, we split our dataset into 70% training and 30% testing. We also included a Shuffle parameter data to address the imbalance of safe/unsafe.
* Explanation of model choice, including limitations and benefits.
  + We changed from supervised machine learning models to a deep neural network.
  + The main benefits of neural network are that it is flexible and can be used for both regression and classification problems and works best with lot of data points.
  + One problem of neural network is that it is prone to overfitting. Adding too much nodes and hidden layer will overtrain our model. Additionally, is is computationally very expensive and time consuming to train models.
* Description of how they have trained the model thus far, and any additional training that will take place
  + In our first hidden layer we added two to three times the number of nodes as input features. With six input features, we put 15 nodes in the first hidden layer and 10 layers in the second hidden layer. Additionally, we added a dropout layer to prevent overfitting. In our hidden layers we also using ‘relu’ activation function and ‘sigmoid’ activation function in our output layer. These activation functions were used because they are commonly used in classification problems. To train the model, we increased the batch size during the fitting process to hyper train our model and then lowered it at our final training. Finally, we lowered the epoch from 100 to 50.
* Description of current accuracy score Additionally, the model obviously addresses the question or problem the team is solving.
  + With the deep neural network model, we have an accuracy of 99.8%. Our model predicts whether the presence of chipotle is determined by crime and population demographic data.