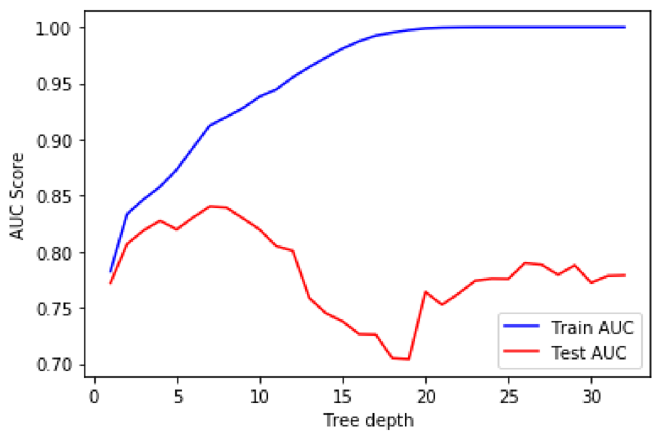
In this analysis, we leverage three different decision tree modeling techniques to determine an efficient and precise way to predict whether an individual would default on their loan, as we all as the amount that would be lost due to default. The three models tested were traditional decision trees, random forest, and gradient boosting. In order to determine if the models under review were accurate, statistics were calculated, such as accuracy metrics, as well as the area under the curve (AUC). These metrics provided a benchmark that would allow for comparison when seeing which modeling technique was the most effective. Also, the root means squared error (RMSE) was calculated in order to determine how accurately the loan default amount was predicted. In regards to classifying whether an individual would default on their loan, It was observed that the random forest decision tree resulted in the most accurate outputs when comparing the AUC test results across all three models. In regards to predicting loan default amount, the gradient boosting model results in the lowest error; however, the random forest was extremely close.

Bingo Bonus:
In order to make sure that the model was optimized, I tuned the hyperparameters of the models in order to make sure we were predicting as accurately as possible. For the max_depth parameter, which decides how 'deep' each tree is built, I leveraged a for loop to calculate the AUC in different depth scenarios and plotted the results to find where the most optimal AUC would be found. The visual made it very easy to see where the mode became overfitted. Below is a screenshot to show an example of the plot.



I also tuned the parameter, min_samples_leaf, which determines the minimum amount of records needed at each leaf node. Utilizing the same technique as max_depth, the for loop allowed me to test different scenarios and to find the optimal output. As min_samples_leaf increases, the AUC decreases.