In-Depth Analysis

**Model Selection and Development**: Due to the nature of my model to try to predict housing prices in a particular place I elected to try two techniques, a Linear Regression Model and a Random Forest Regressor model, and see which model worked best on my dataset.

**Linear Regression**: Using a Linear Regression model to predict bedroom prices using State, Income, and Population, it actually performs pretty terribly, with an R^2 value of 0.261. The data was split into a training and testing set using train\_test\_split at 80% used for training and 20% used for testing. Since the model’s poor performance, I elected to try using a Random Forest Regressor, since the data does not necessarily need to be linearly correlated to be able to predict accurately. It is also quite robust and one of the more accurate algorithms out there.

**Random Forest Regressor:** I tried using the Random Forest Regressor untuned, with the same training/testing split and the model actually performed even worse than the Linear Regression with an R^2 square score of 0.149, that is, accurate only less than 15% of the time. It also had a mean absolute error of 2545.414666115432. However, after tuning the hyperparameters greatly improved both the accuracy and error values of the model.

**Tuned Random Forest Regressor:** The hyperparameters used to train the dataset were as follows:

* max\_depth: A list from 1-20
* N\_estimators: a list from 1-20
* Max\_features: a list from 1-3

Using GridSearchCV with 3 partitions, the model was tuned with the training set, and had the following calibrations:

Tuned Random Forest Regressor Parameters: {'max\_depth': 7, 'max\_features': 1, 'n\_estimators': 18}

And a Best Score of : 0.5588587564224368

On the testing set, the model had a R^2 value of 0.5352073826241069 and a mean absolute error value of 2287.140429461093. By tuning the model, the accuracy improved more than three-fold and had a reduced error.

**Possible biases:** The data available is more available in more popular, urban counties than other rural counties across the US.