Part B)
Our models resulted in RandomForest (RF) classification performing comparably to GradientBoost (GB), while the DecisionTree (DT) model performed the worst. Knowing that RF is a well-suited algorithm for handling noisy data and that our models were trained on 55 different measures, we were expecting RF to give the best performance. Also, many of the measures used in training our models were Null/zero for countries not included in the 9 found in our database (i.e. WHB Health measures for Iceland were not included in our database, but the DevelopmentIndex measure was). This left us with a balanced number of IncomeGroup (our target variable) entries, but only a handful of meaningful measures for the other countries. Even though our dataset was missing many values, the fact that it had a balanced number of IncomeGroup entries allowed the GradientBoost model to perform surprisingly well. In regards to Precision, Accuracy, and Recall scores, had we collected data for all measures and countries the GB model would have surely beat out the RF & DT models.

To add; we decided to train our models on the data collected for all countries (regardless of the many missing values, which we replaced with 0). To nobody's surprise, training our models on only 9 countries resulted in 0% accuracy/precision for each of the three models, as we were trying to predict the IncomeGroup of a country, which has a 1-1 relationship.

In summary, training our models on the DevelopmentIndex measures for each country resulted in surprisingly good precision/accuracy/recall scores, but there is still huge room for improvement. Collecting the Health, Population, and Education measures for all countries will be the best course of action to improve our database.