Consider the Boston housing prices data set.  Inspect it, and create one record (one row) and show it as the first part of your answer.  Leave the predicted variable Y\_i as unknown (property price in 1000$). In this context, answer the below questions integrating your answers like an essay (tell a story):

[Boston Housing Price Data (Links to an external site.)](https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html)

1) What machine learning algorithms would you use to predict the property price (home price Y\_i)?

2) Suppose we convert the home Price into a multiclass category.  how would you achieve that using binning (converting numbers to classes)?  what algorithms would you use now?

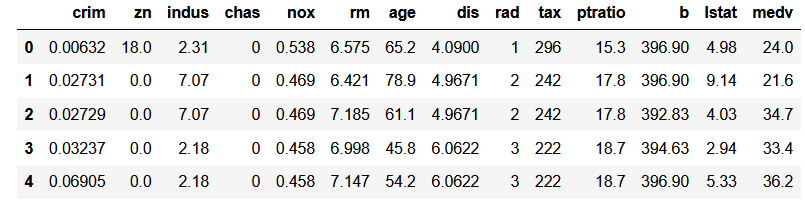
Specifically in the context of predicting the answers for this one record, please answer the below questions:

3) What is bagging and how does it work?

4) How does boosting work, and is it better than bagging? Why?

5) Briefly explain what these algorithms are and how they work: AdaBoost and XGBoost

The Boston Housing Prices data set contains 506 rows (records) with 14 columns (features).



The target variable for the noted dataset is “medv” which stands for **Median value of owner-occupied homes in $1000's.**

For this particular problem when we want to predict a price (continuous variables) based on multiple features a multiple regression is the best way to solve this problem. The regression itself can be done by multiple algorithms such as: multiclass linear regression, regression trees, regression assemble models (Random Forest for Regression).

Another way to solve this prediction problem could be by converting the “medv” target variable into a multi-category variable by splitting the data in different buckets (Ex: High, Medium, Low) based on the distribution of such variables and some desired boundaries to establish the side of the bucket (category).