

EAI 6020 Applications of Artificial Intelligence

# Assignment Information

Name: Dickson Wanjau

Module: Data Science And Machine Learning Platforms

# Instructions

Replace all text highlighted in yellow for your own submission. Use this word document to describe how you complete each step in the problems below. Show all commands, screens, pages, and options. Points are deducted for missing steps.

When complete, upload your MS Word document named: EAI6010\_LastNameFirstNameWeekX.docx to the course website in your folder for the Module 5 assignment. Do NOT include screenshots as separate files.

If you have issues with the upload, please notify your instructor. If you are raising issues that might be of interest to your colleagues in the class, please use the assignment thread on the discussion board. The discussion board is your friend.

# Activity

Problem 1:

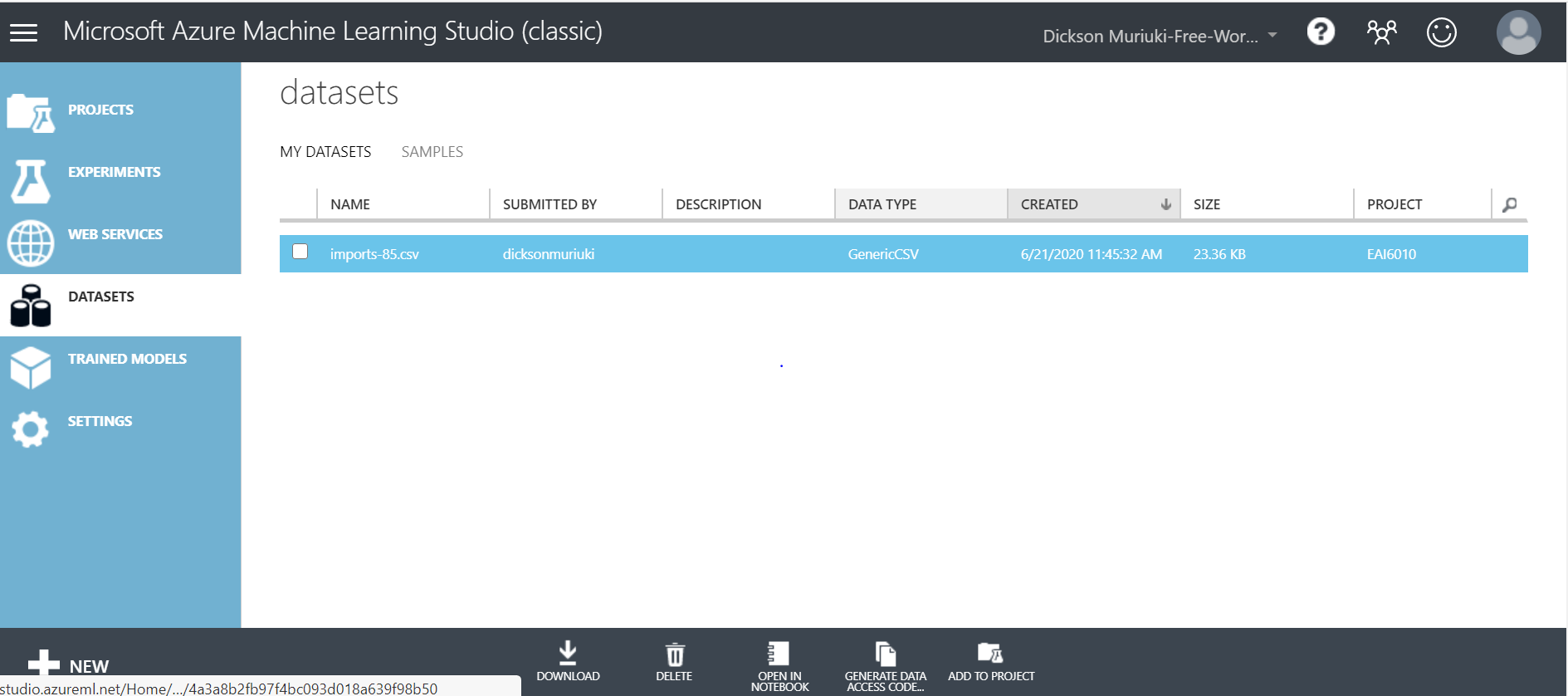
1. Use regression techniques on Azure ML Studio to create a model for predicting car prices.
2. Start with the import-85.csv file containing car characteristics and prices, extract four features you consider most relevant for setting the price of an automobile.
3. Create models using three of existing regression algorithms and report which of them performs the best with the features you selected.
4. Please be aware that Neural Network Regression model can take several hours to train and we recommend that you do not use it for this activity.
5. Upon completion, remove all Azure resources that you created.

# Results

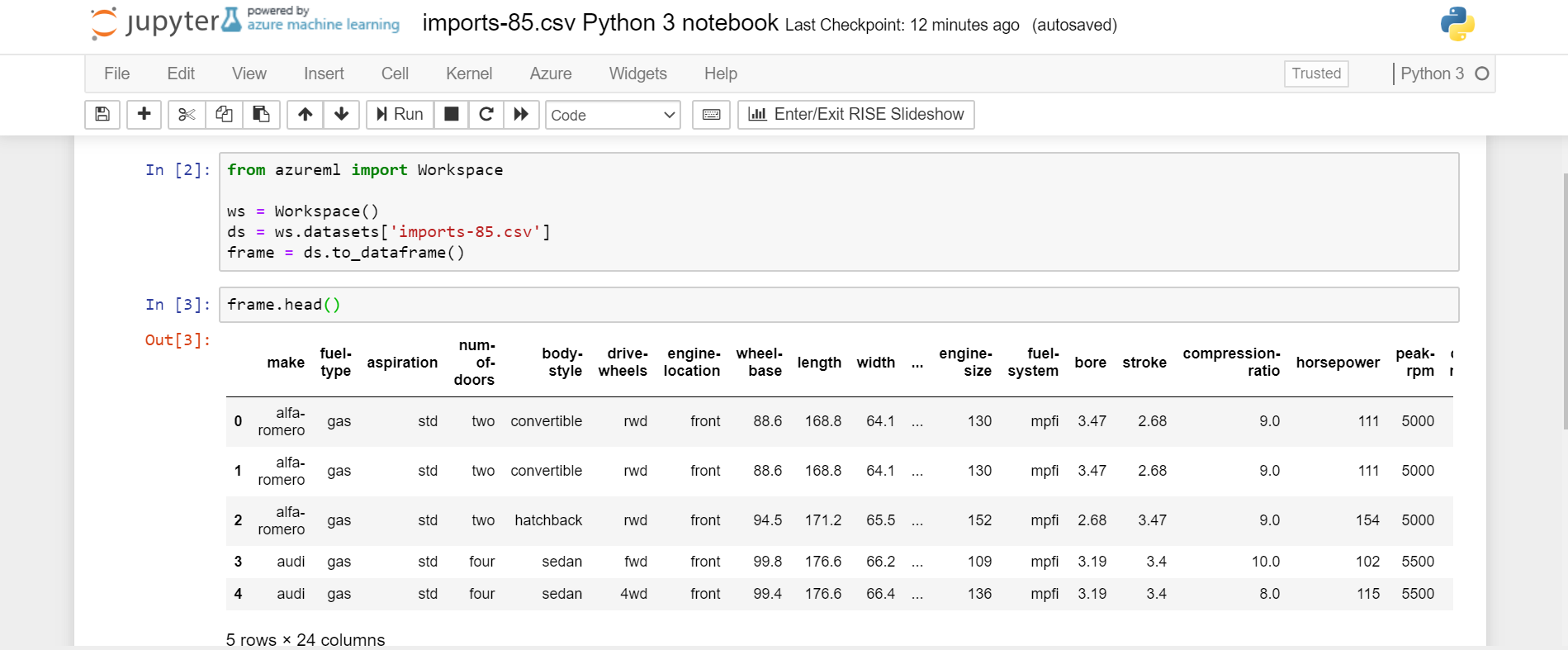
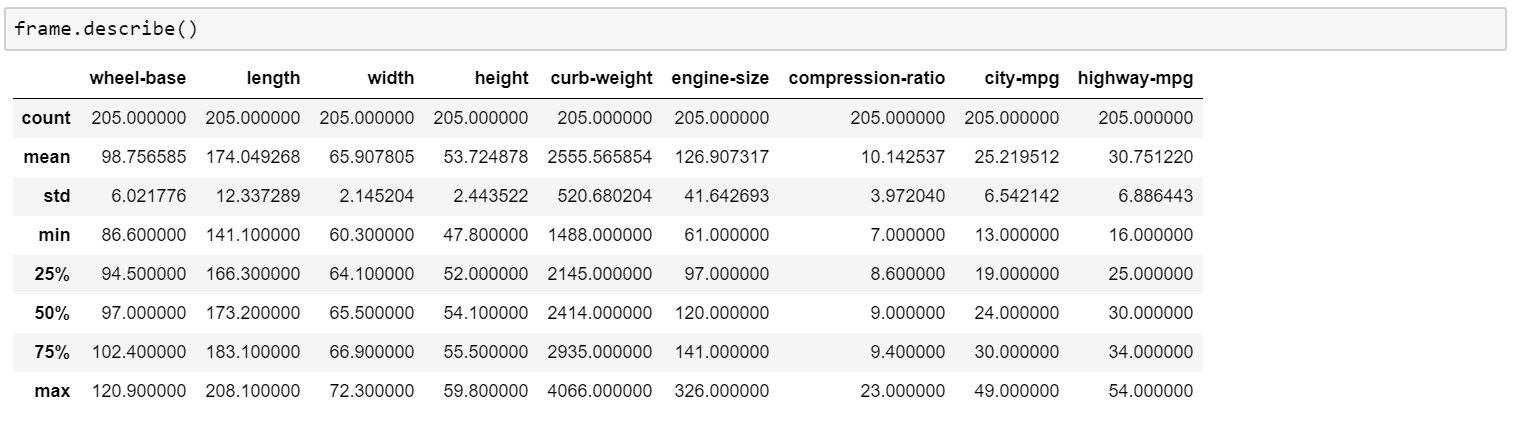
Describe the steps, paste screen captures, and answer any questions for each problem in the activity here.

Problem 1:

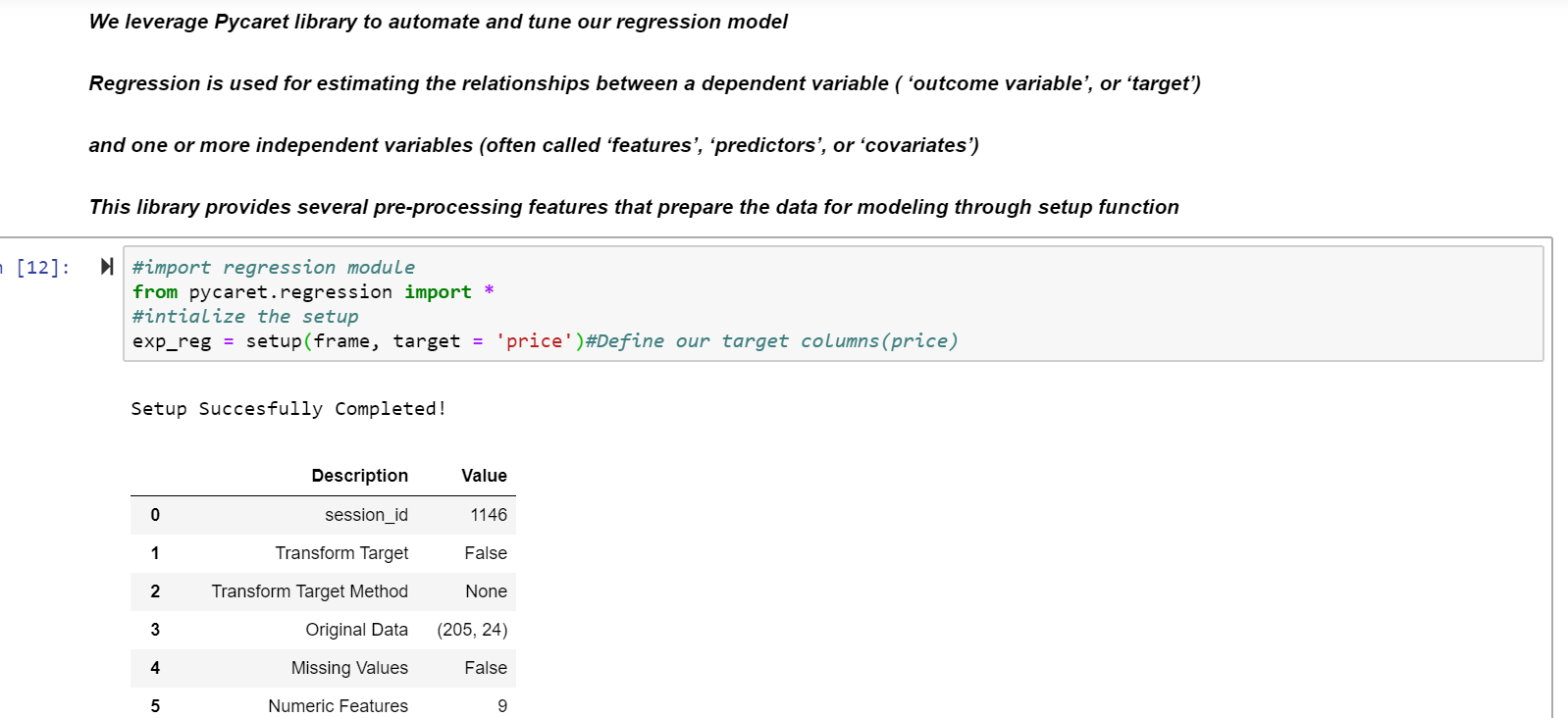
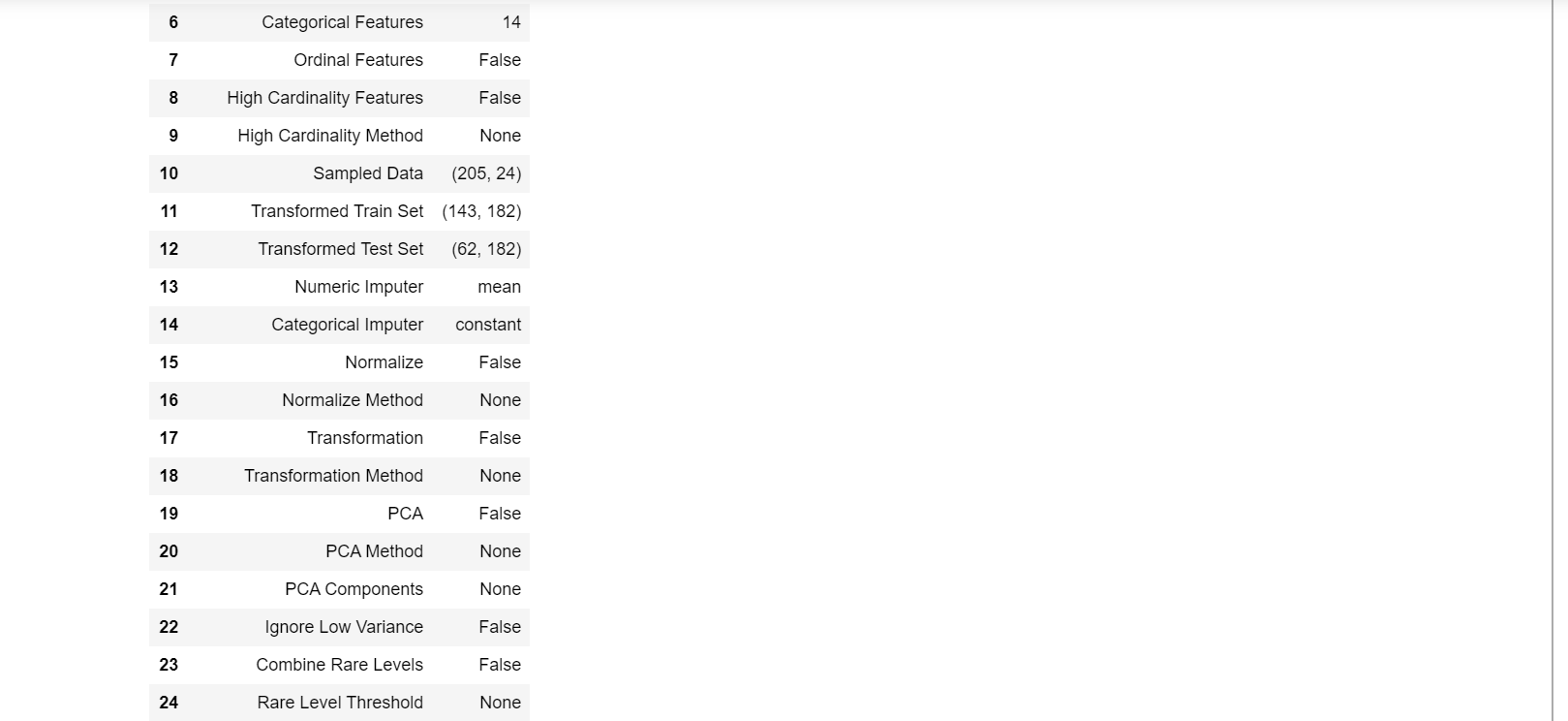
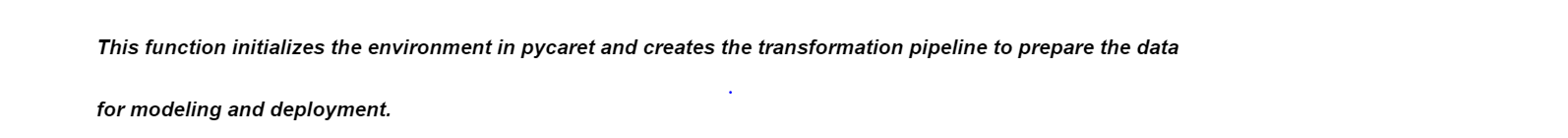
Use regression techniques on Azure ML Studio to create a model for predicting car prices.



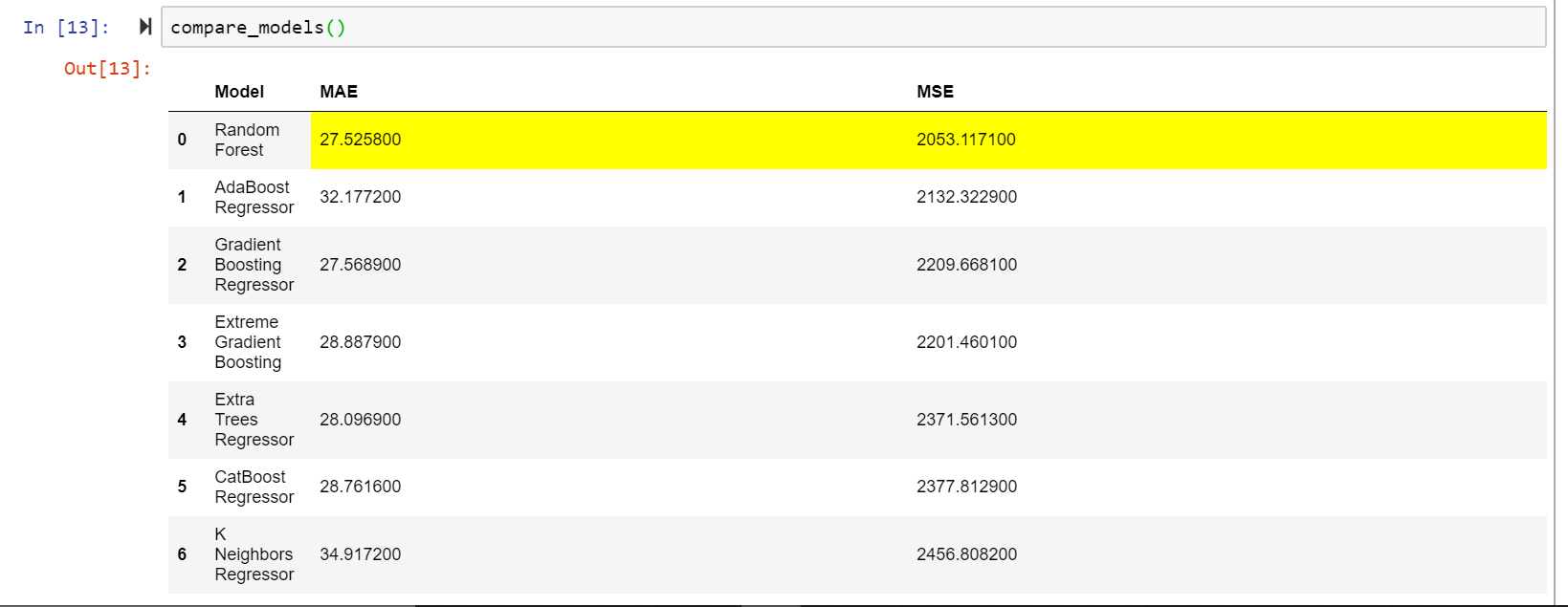
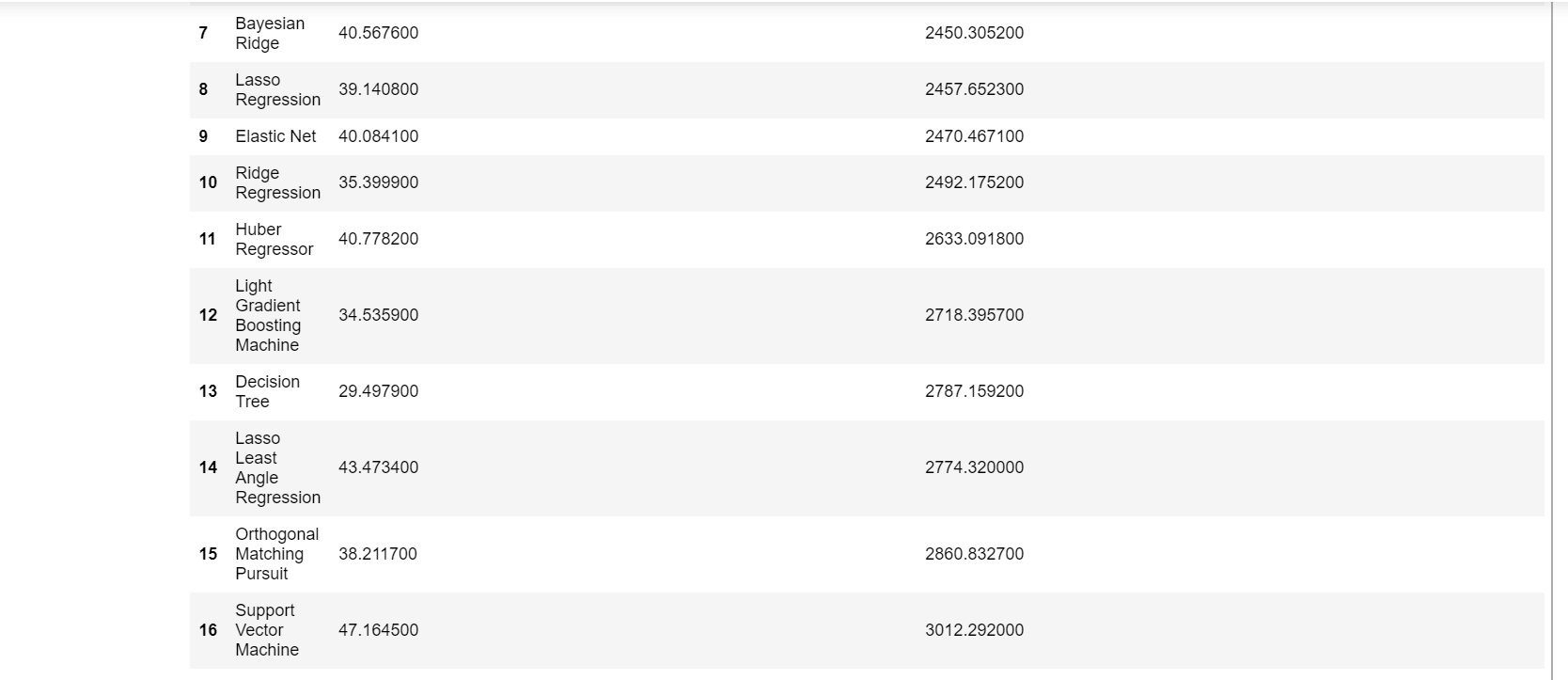
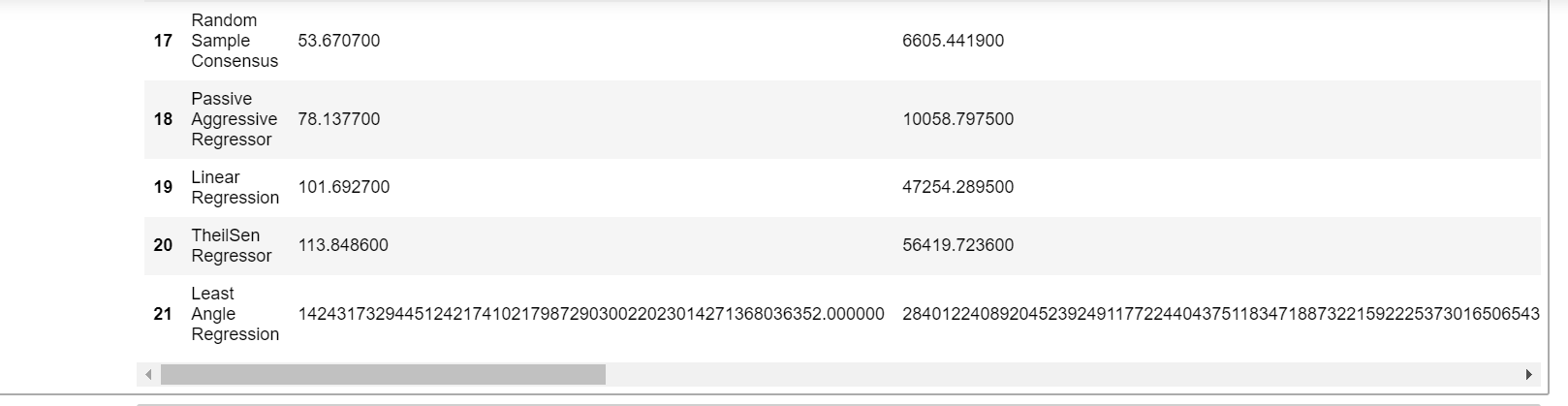
b) Start with the import-85.csv file containing car characteristics and prices, extract four features you consider most relevant for setting the price of an automobile.

c) Create models using three of existing regression algorithms and report which of them performs the best with the features you selected.

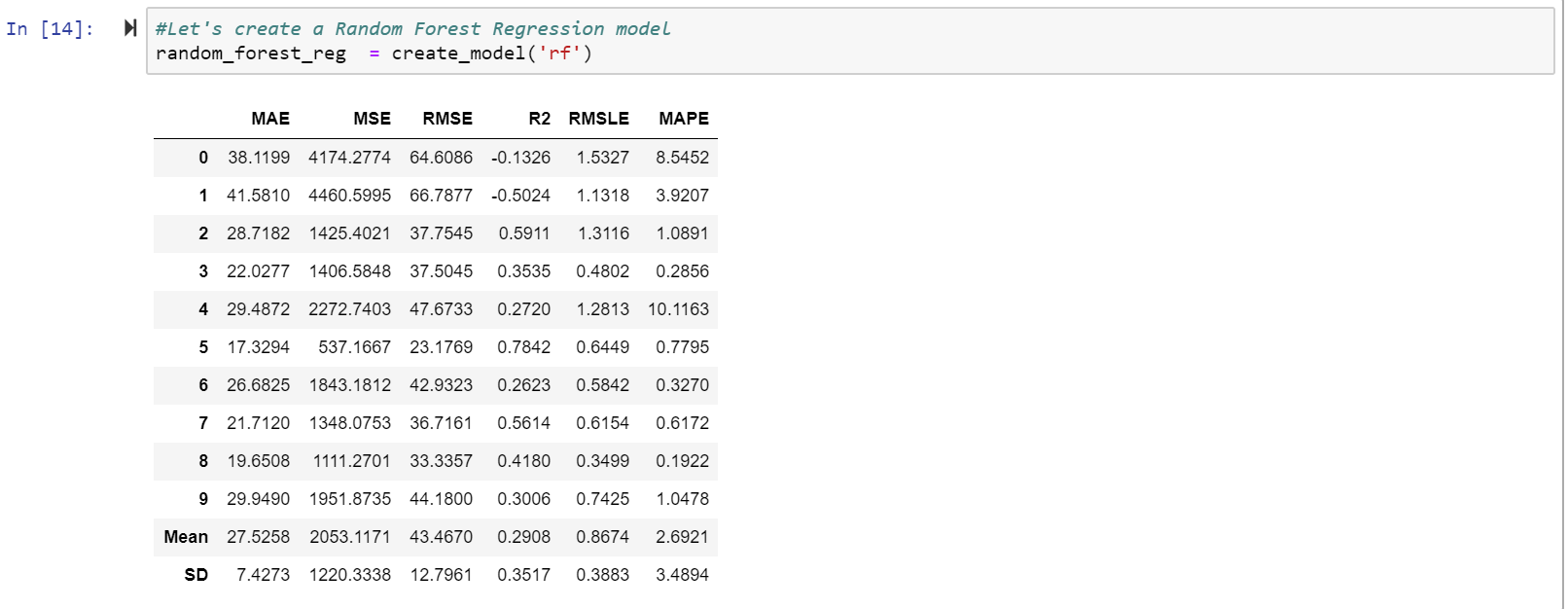
   

Next, we leverage the compare\_models() function which uses all models in the model library and scores them using K-fold Cross Validation. The output prints a score grid that shows MAE, MSE, RMSE, R2, RMSLE and MAPE by fold (default CV = 10 Folds) of all the available models in model library. We can always choose the number of folds that we want to perform using the fold parameter inside the function.

From our output, we note that the Random Forest model is the best performing regression model with a Mean Absolute Error value of 27.5 and a Mean Square Error value of 2053.12.

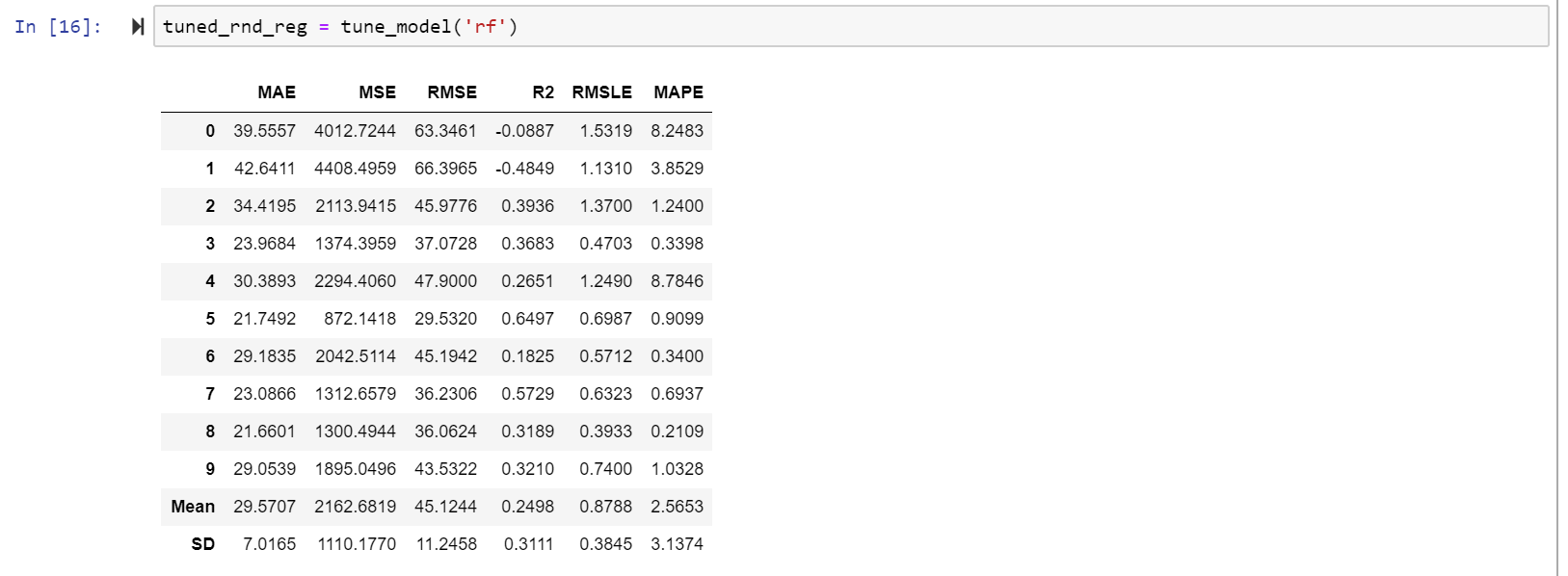
Next, create 3 regression models, based on the results returned and observe their results. We’ll create a Random Forest regression model, AdaBoost regression model and Gradient Boost regression model.



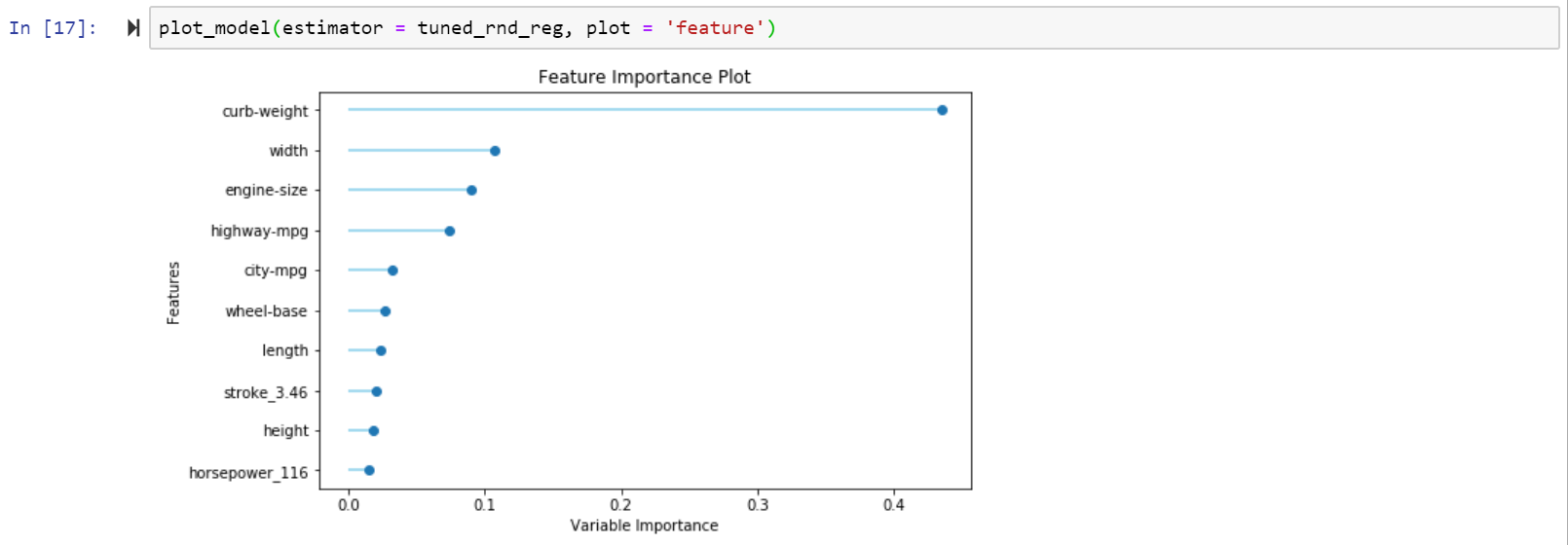
The score grid returns a table containing the scores of the model across the k-folds (10-folds). Scoring metrics used are MAE, MSE, RMSE, R2, RMSLE and MAPE.

The mean and standard deviation of the scores across the folds are also returned.

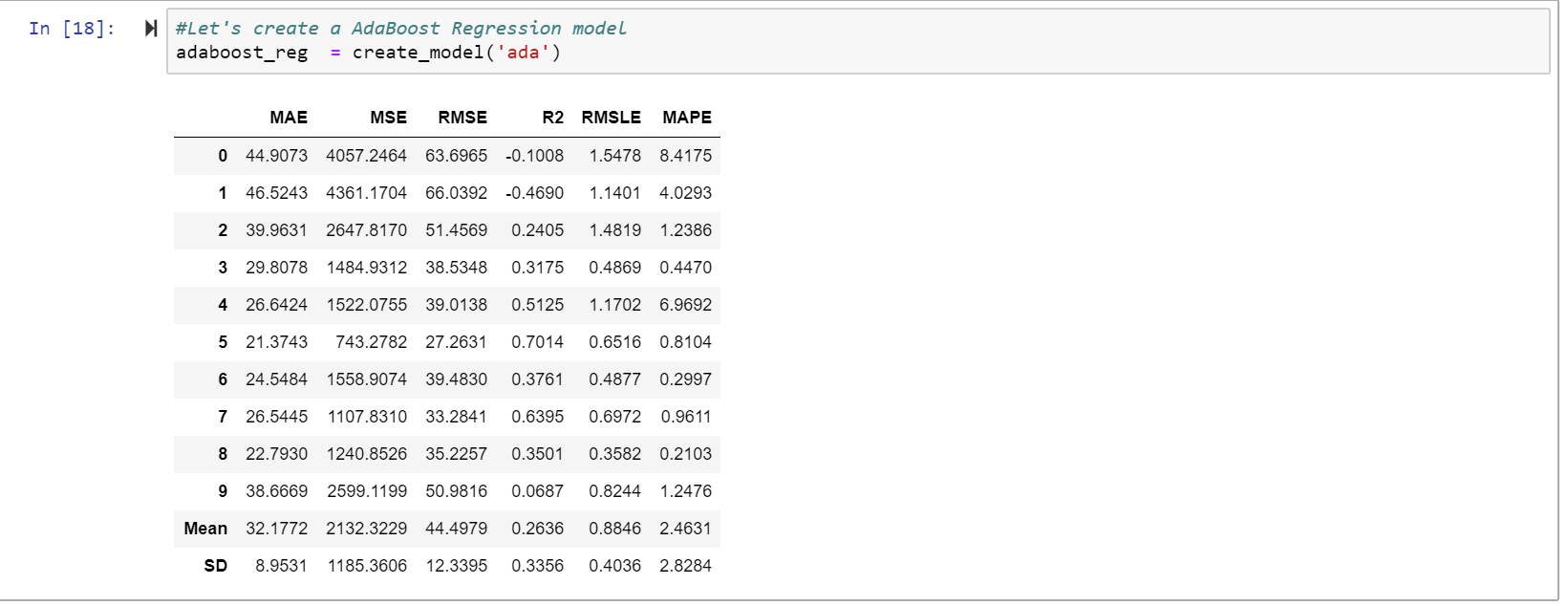
Let’s now tune the hyperparameters of the and score the model using K-fold Cross Validation. We achieve this using the tune\_model () function which has a default 10 fold cross validation.

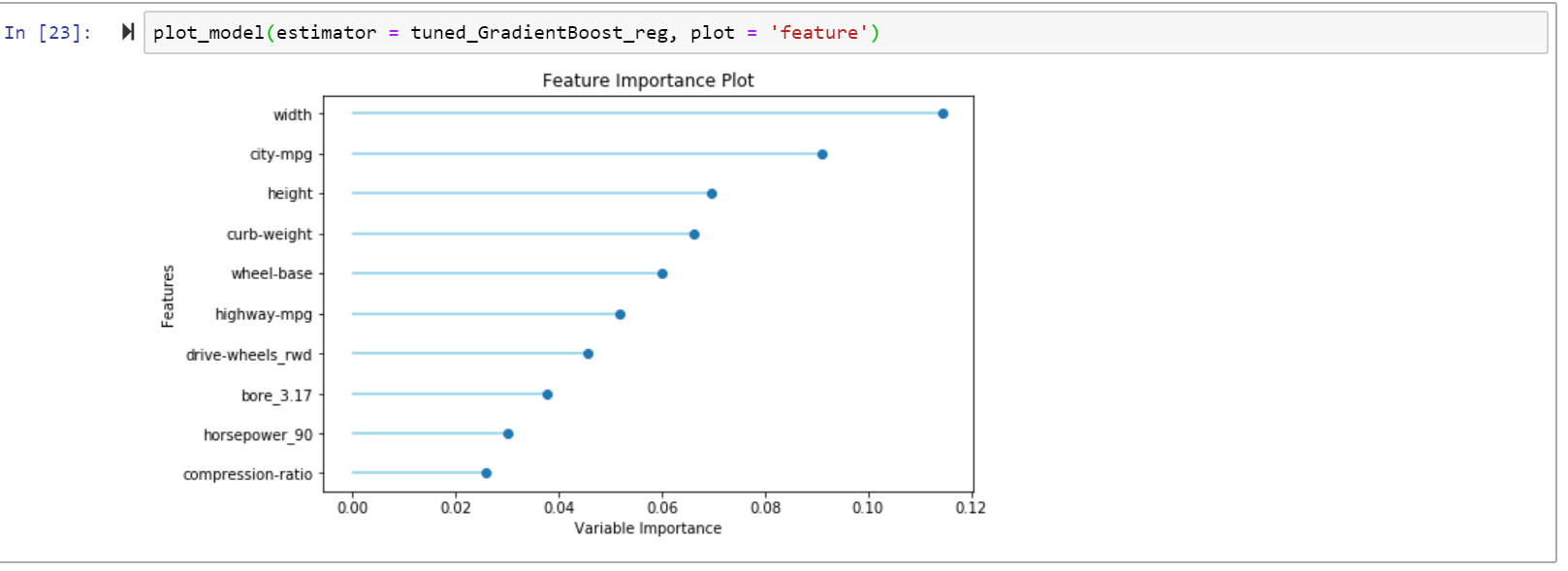
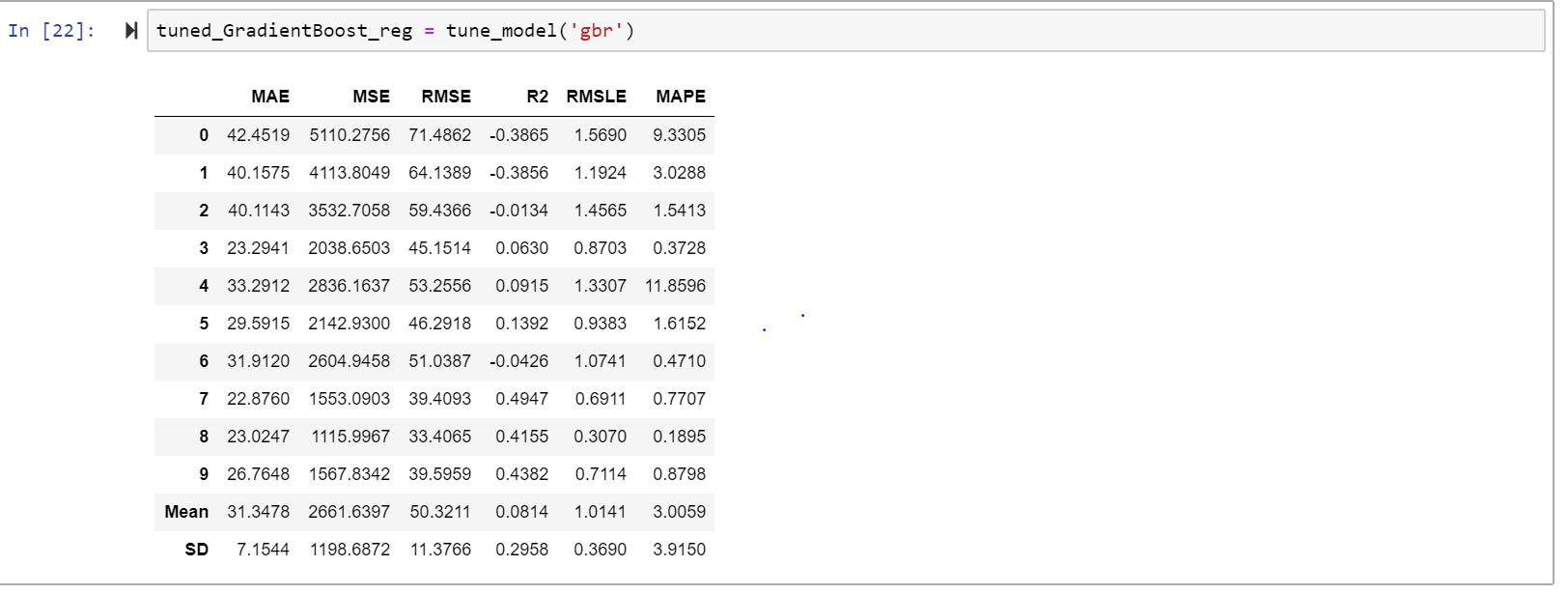
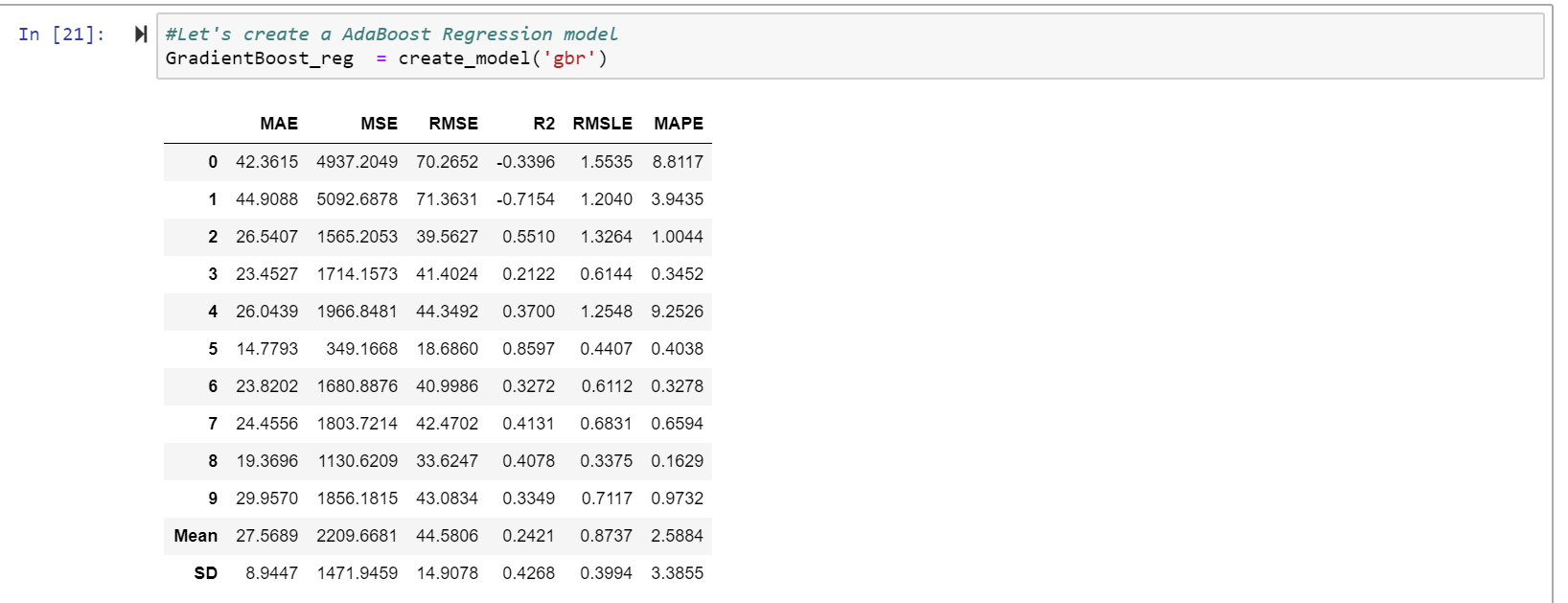
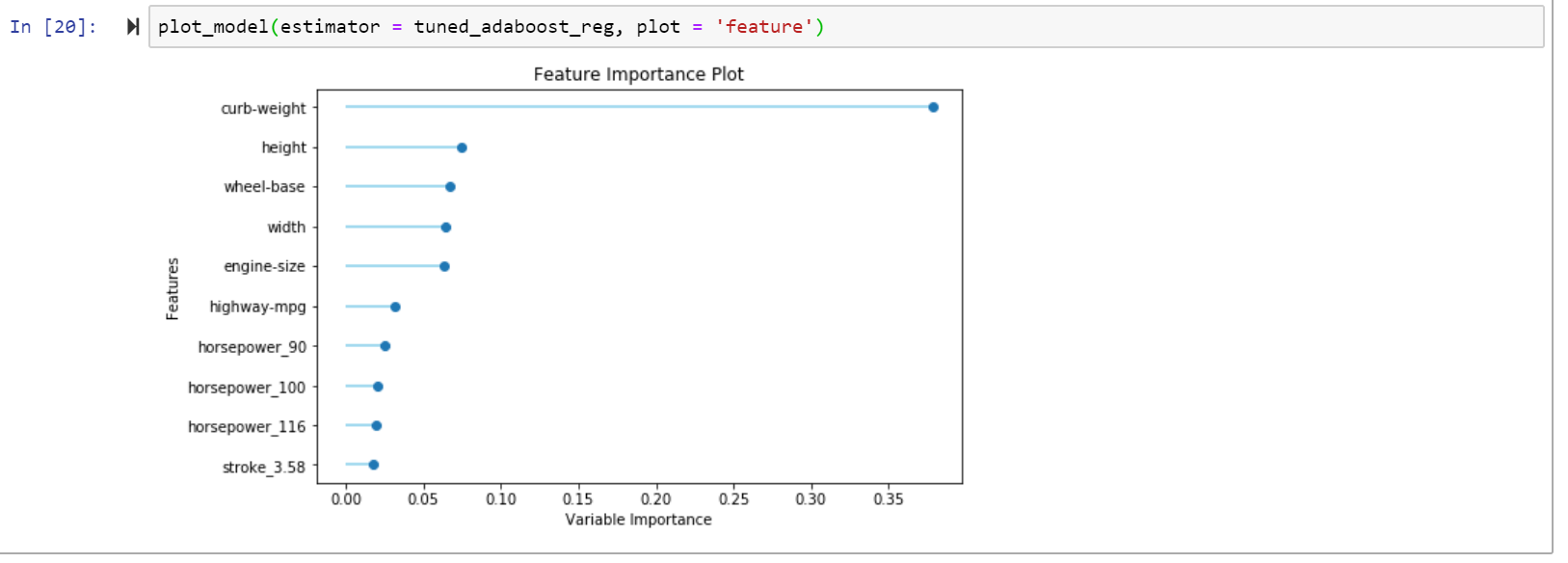
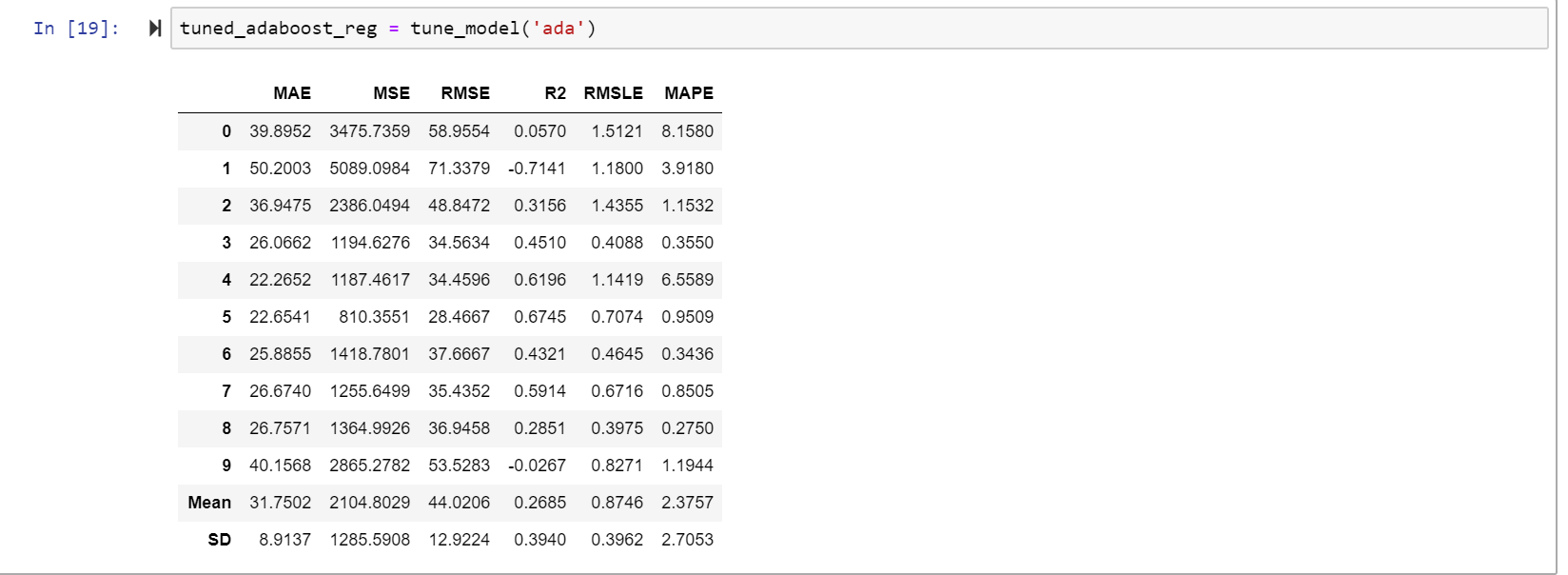


Finally, we plot the feature importance plot to in order to get a quick understanding of what features actually matter.



Let’s repeat the above procedure for the AdaBoost regression model and Gradient Boost regression model respectively.





e) Upon completion, remove all Azure resources that you created.

