PETRONAS

Regression

Citizen Analytics – An Initiative by Data Science Team

START ▶

Learning Objectives

By the end of this module, you will be able to:

01 Define regression and how it works

02 Describe common regression algorithms

03 Learn how to perform regression in Azure ML Studio

PETRONAS

Content

Linear Regression

PETRONAS

What is Linear Regression?

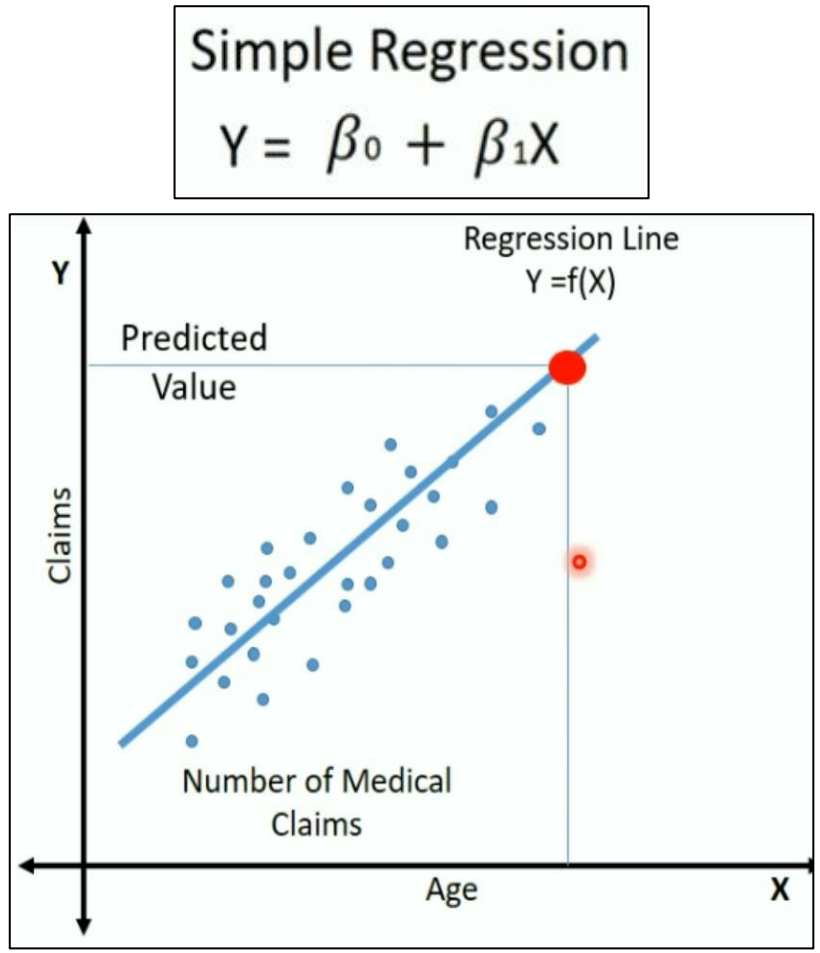
- Use Linear Regression when you want to understand something which is continually varying.
- Linear regression attempts to fit lines to your data, by minimizing an error function.
- Determine the relationship between the dependent variable (response) and one or more independent variables - predictors.
- Predictors and response variables are continuous variables.

Examples of continuous variables:

- Temperature of a boiler
- Flow rates
- Numbers of people expected
- Total spend

Simple Regression

- Y = mX + C
- m is gradient, C is y-intercept
- The line is fit to best match the data, based on error function.
- Given a new X, new estimate of Y can be made after parameters fitted.



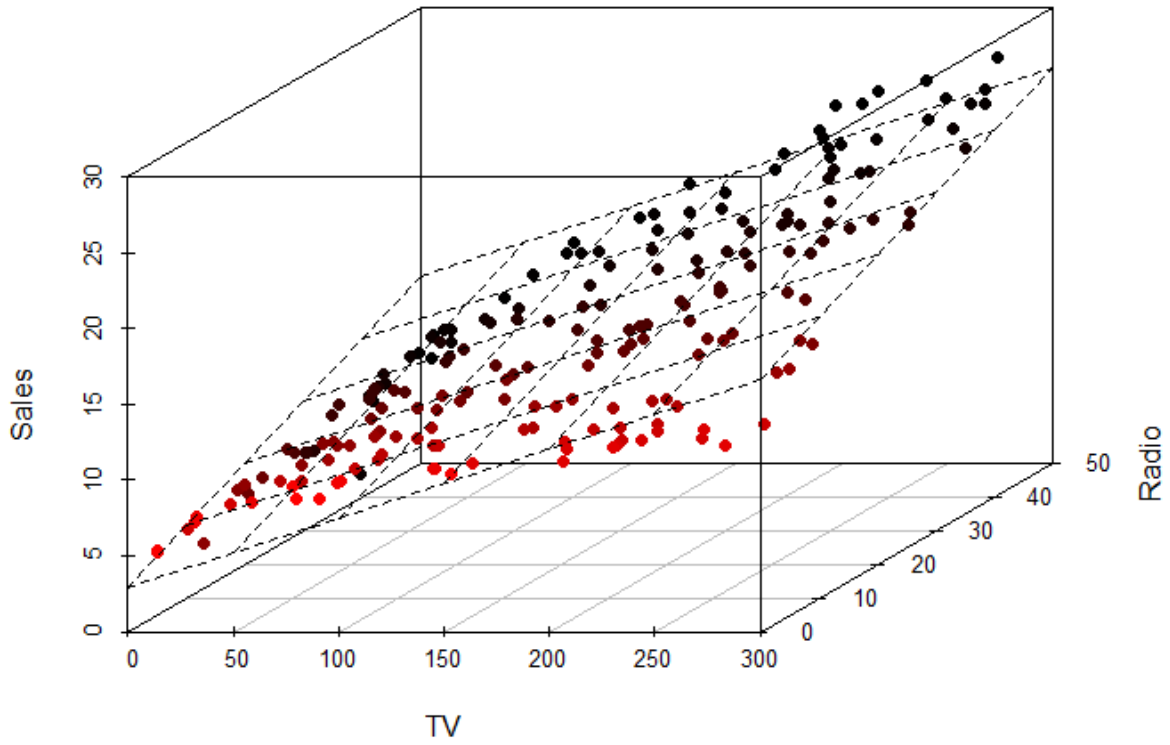
Source: Simple Linear Regression. https://www.tech-quantum.com/classification-logistic-regression/

Multivariate Linear Regression

- Y = mX + C but in higher dimensions.
- Gradient and y-intercept parameters for every predictor.
- Hyperplane fit through the data in higher dimensional space.
- Given a new set of X_n, the new estimate of Y can be made after parameters fitted.

multivariate linear regression.

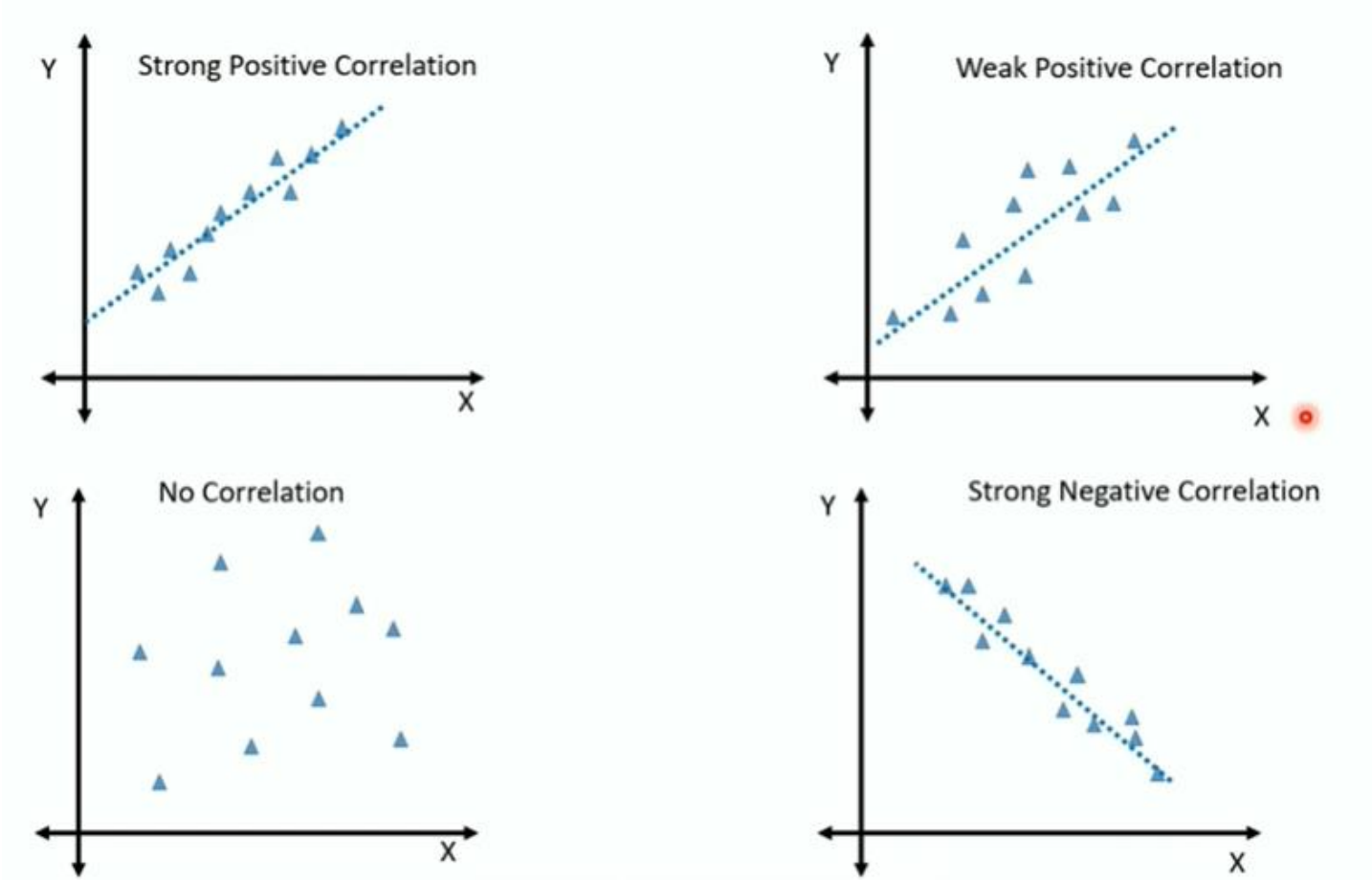
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 ..... + \beta_n X_n$$



Source: Multivariate Linear Regression.
https://stackoverflow.com/questions/26431800/plot-linear-model-in-3d-with-matplotlib

Fitting functions

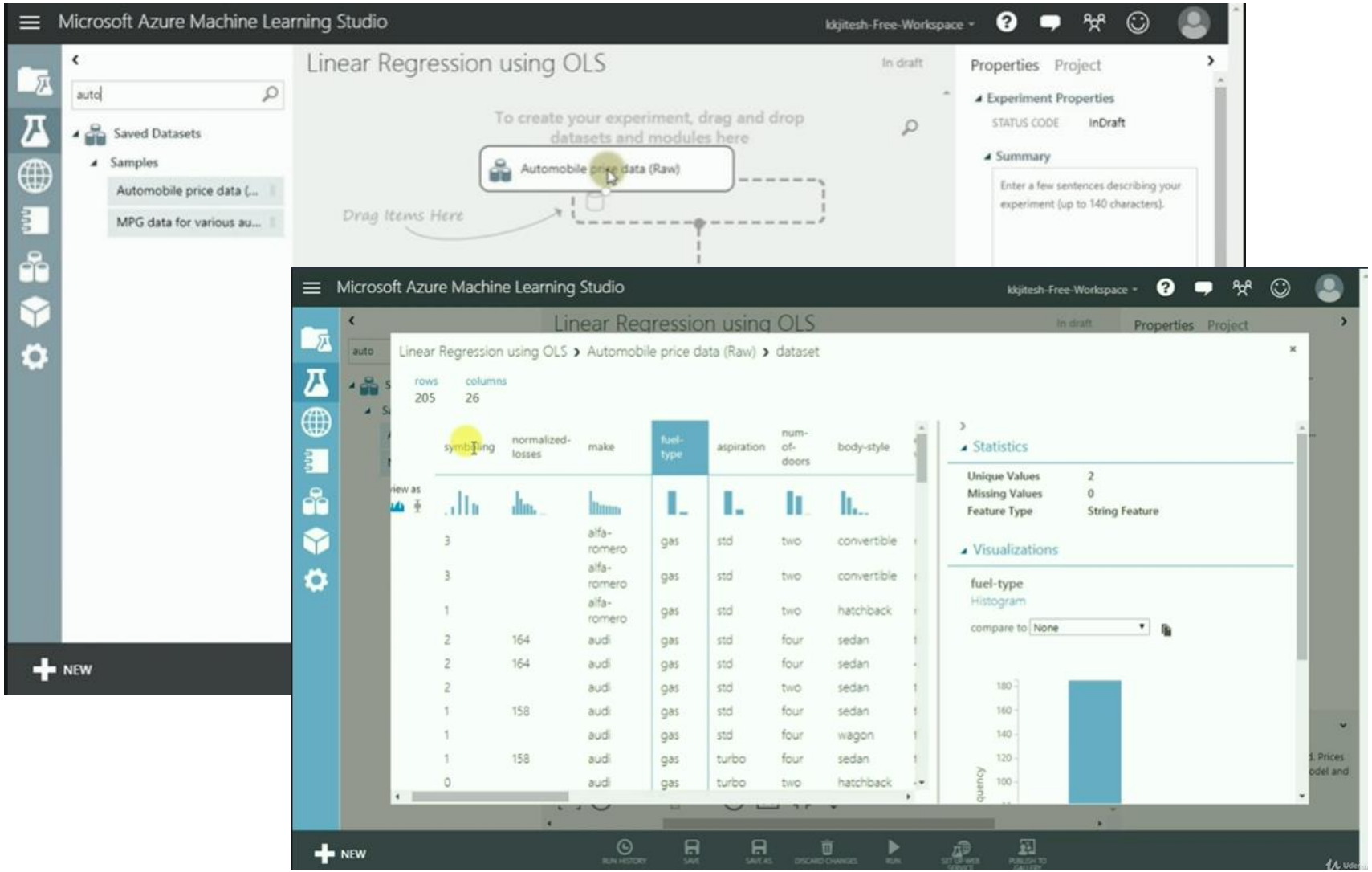
- What is the best line through your data?
- Controlled by fitting functions:
  - Ordinary least square in Y
    - $\Sigma_{data}(Y_{data}\text{-}Y_{line})^2$
  - Mean Absolute Error in Y
    - $\Sigma_{data}\frac{1}{n}|Y_{data}\text{-}Y_{line}|$
  - Root Mean Sq Error in Y
    - $\sqrt{\frac{1}{n}\Sigma_{data}(Y_{data} - Y_{line})2}$



Source: Correlation and Linear Regression. https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Correlation-Regression/BS704_Correlation-Regression_print.html

Linear Regression in Azure ML

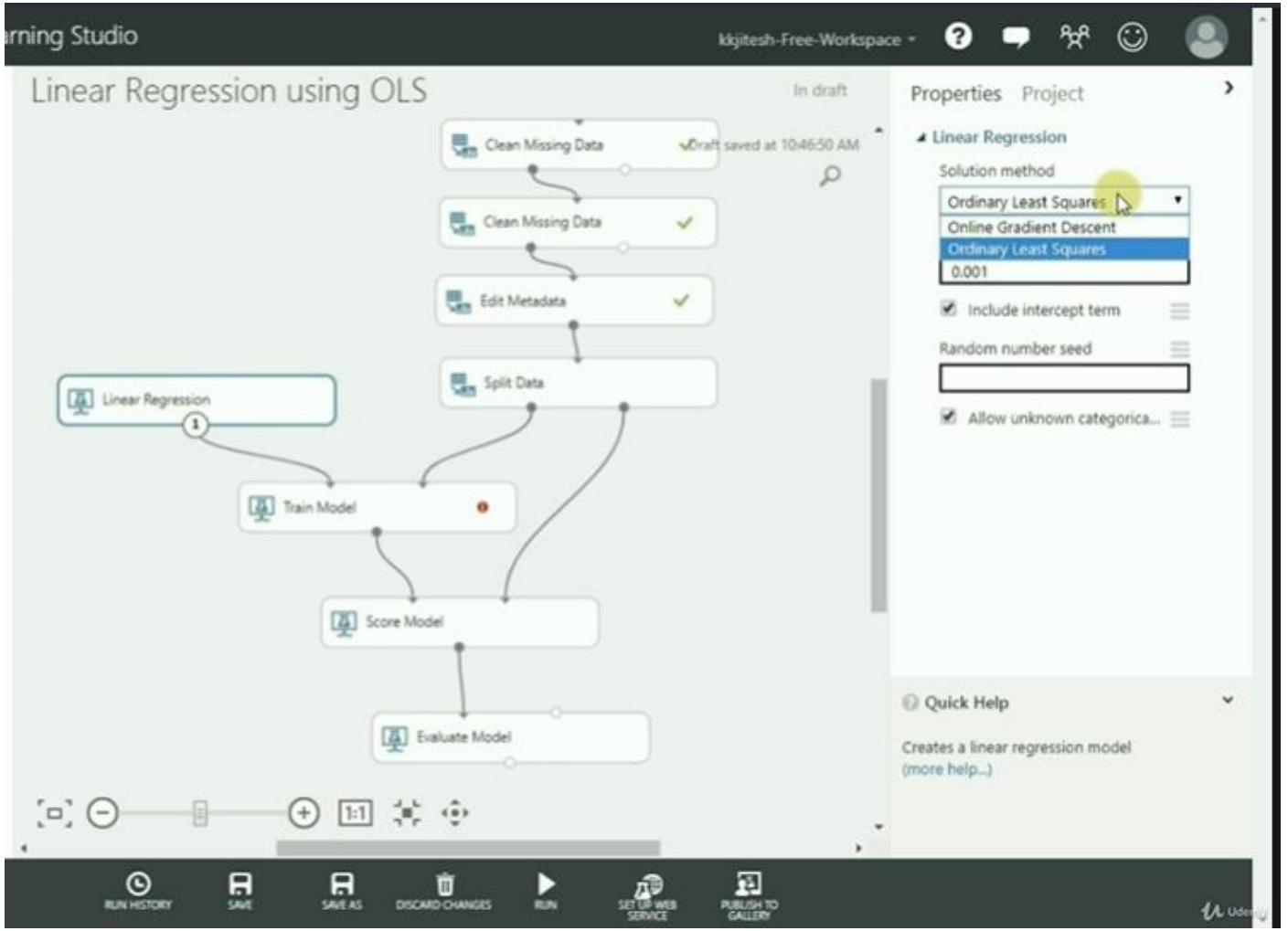
1. Load the data.
2. Right click on the data module and select Visualize.



(Cont…)

Linear Regression in Azure ML

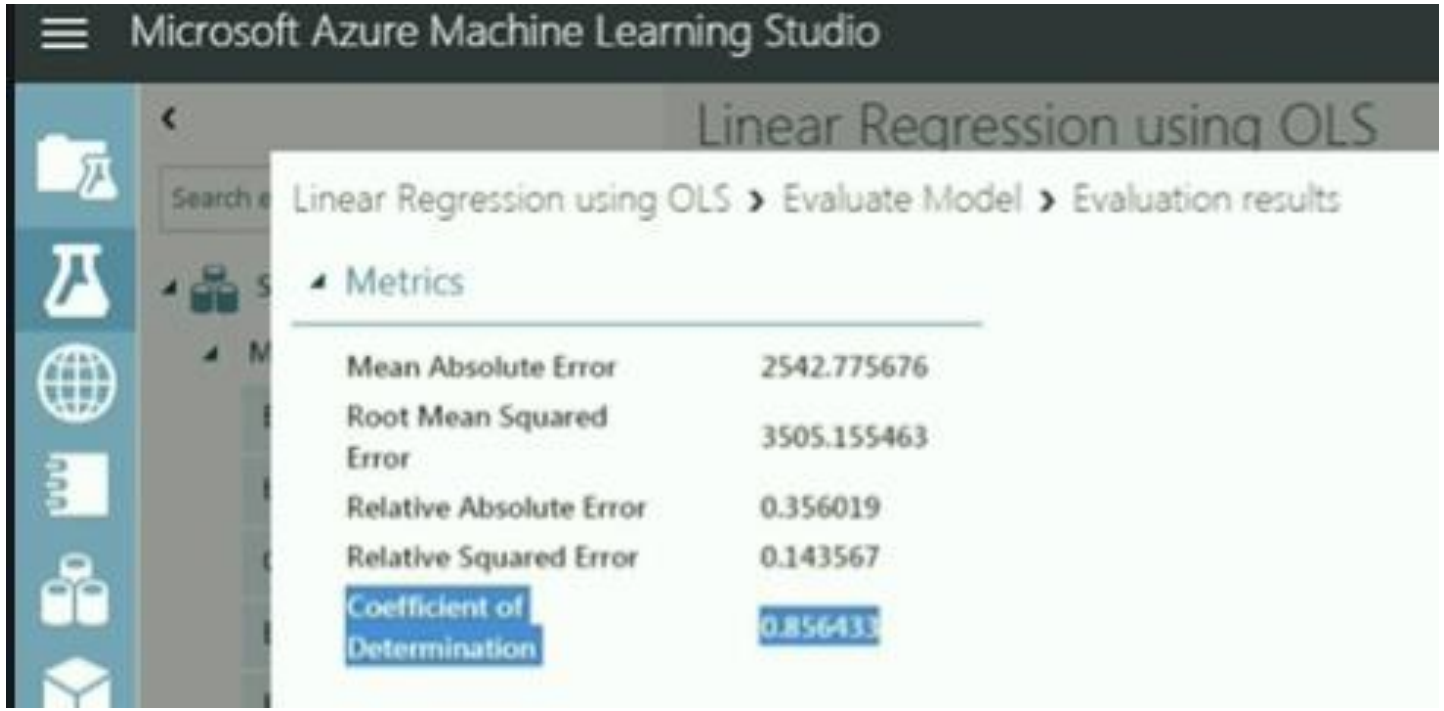
3. Add data cleansing steps and split data into test and training parts.
4. Add Linear Regression model, choose Ordinary Least Squares for error function to minimize over data.
5. Add Train Module block to fit the parameters based on training data.
6. Score Module and Evaluate Model blocks allow evaluation of the quality of fit of data, and performance against the test dataset.



(Cont…)

Linear Regression in Azure ML

7. Check the quality of fit metrics to evaluate the quality of the model.

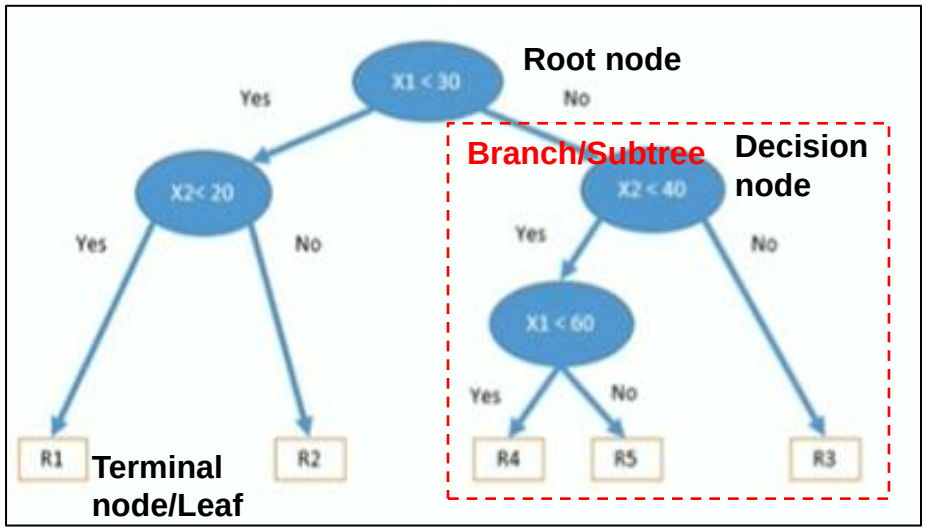


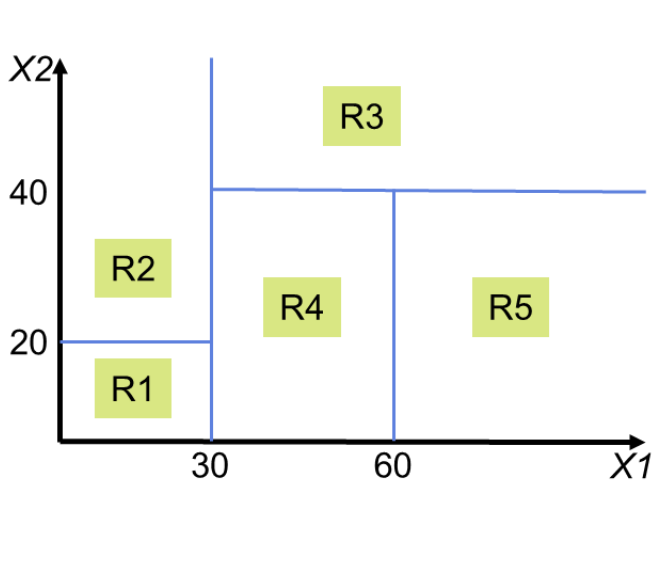
Decision Tree Regression

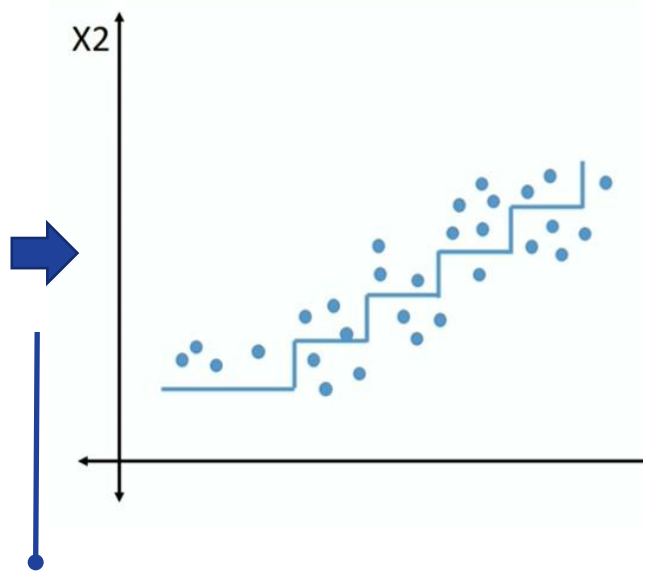
PETRONAS

What is Decision Tree Regression?

- Allows a decision tree approach to be taken with predictors.
- Dependent variable (response) still continuous variable.
- Categorical (only takes certain values) data can be used as a predictor.







To Fit Y:

- Output of each region is mean of region.
- Every leaf may have regression line for points contained within that region.

Source: The parts of Power BI. https://docs.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview

Decision Tree Regression in AzureML

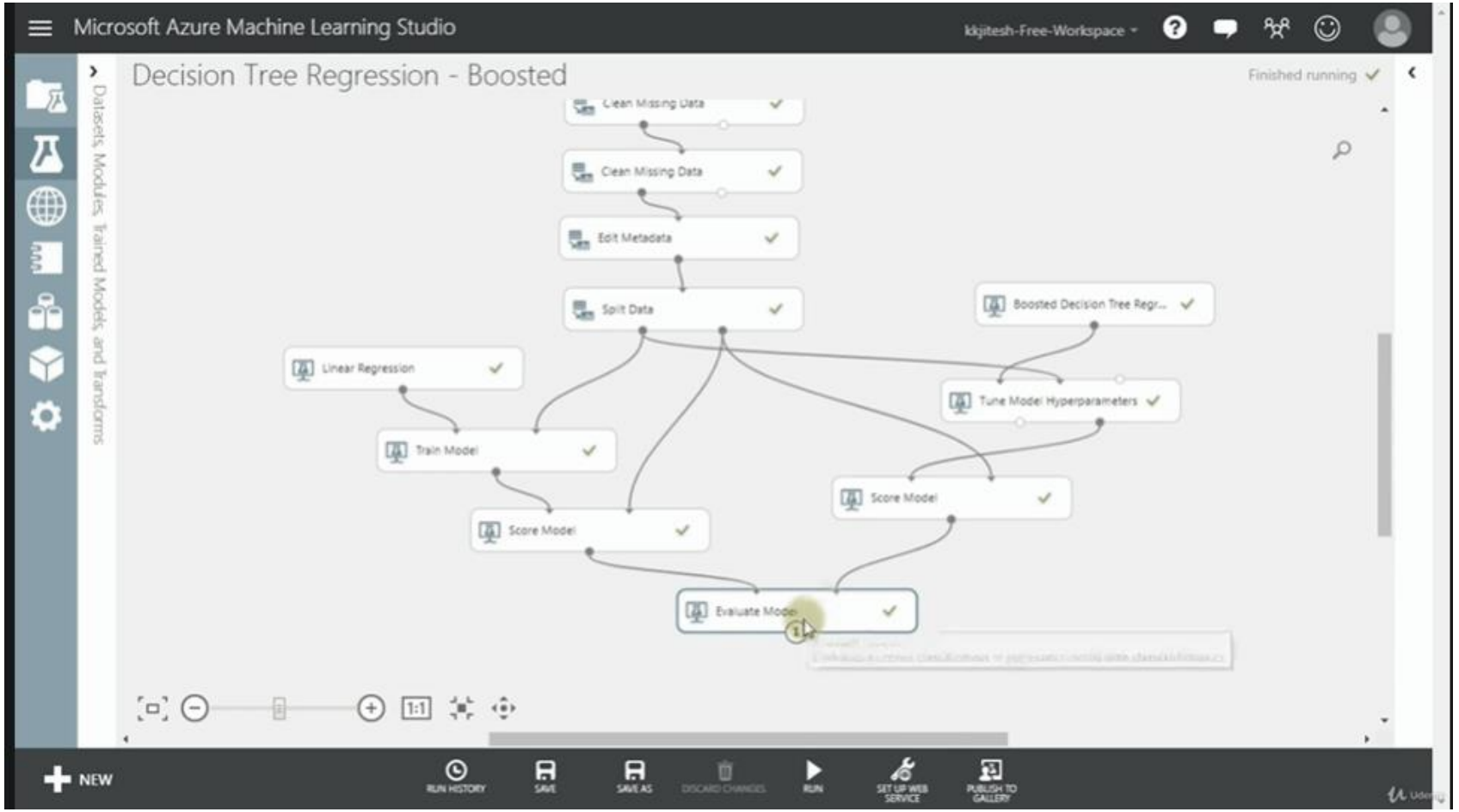
1. Can put Decision Tree Model in same canvas as Linear Regression for comparison.

2. Add Boosted Decision Tree module block.

3. Add Tune model hyperparameters for training model (different from Linear Regression).

4. Add score model block and connect into evaluate model block to compare 2 approaches.

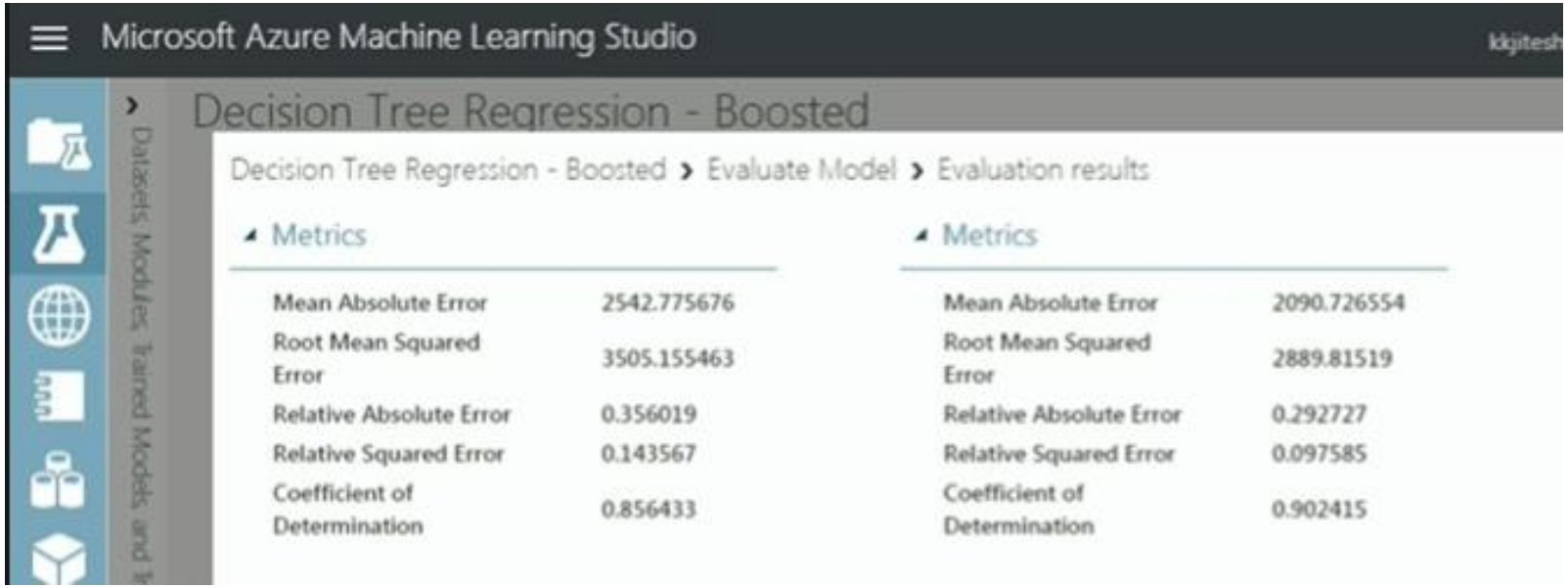
(Cont…)



Decision Tree Regression in AzureML

5. Check quality of fit metrics to ensure model results look good.

6. By running two approaches in the same canvas you can compare 2 models against each other with the Evaluate model block.



Summary

Summary

1

What is regression?

- Regression is a supervised machine learning method to predict a continuously varying output. (e.g. Temperature of a boiler, flow rates, number of people expected, total spend)

2

Different regression algorithms available in Azure ML Studio

- Linear Regression and Decision Tree Regression etc.

3

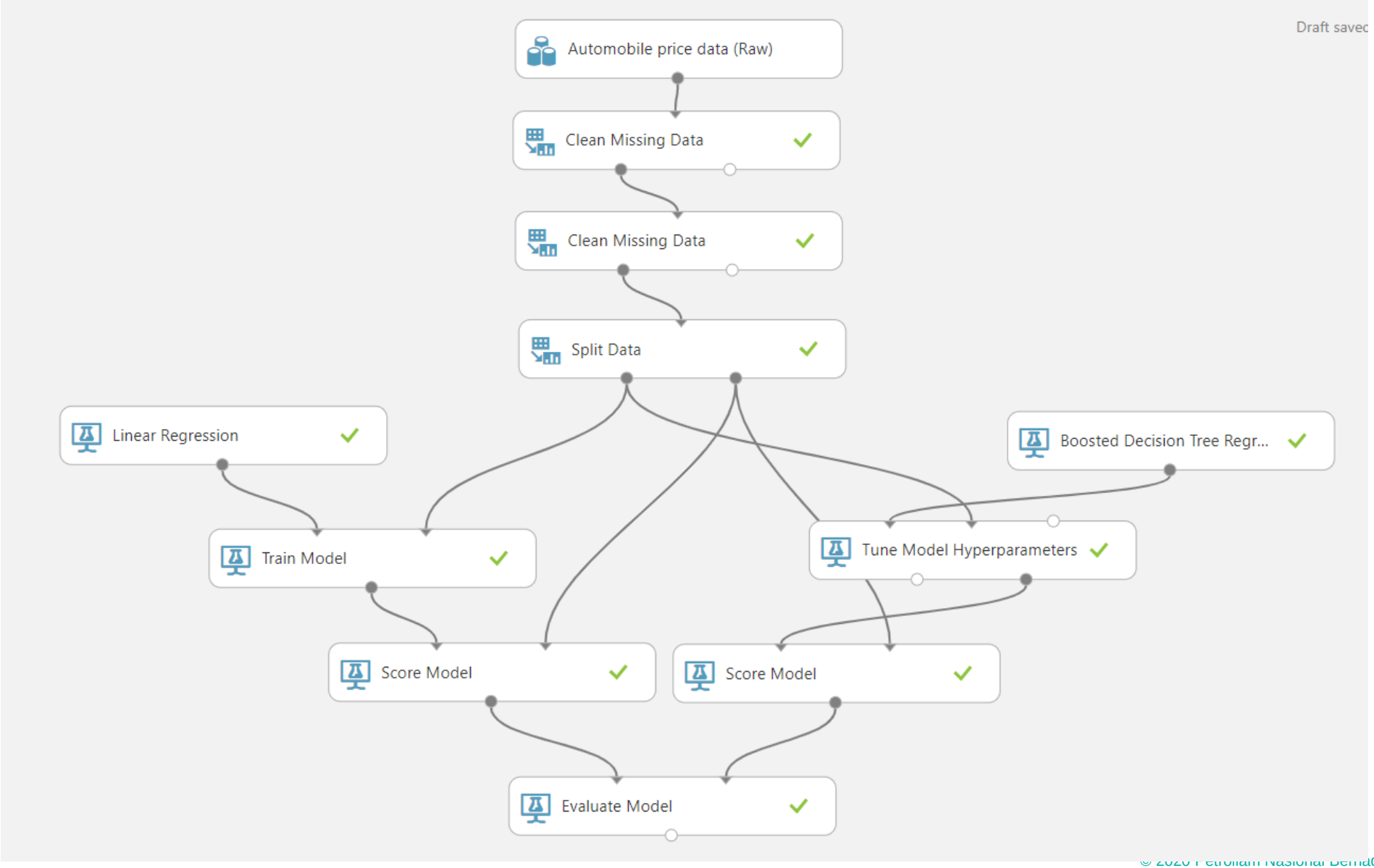
Other facts on regression

- Linear regression can be used to extrapolate (predict outside range used for training) but be careful!
- Regression models can fit the data in different ways, controlled by the error function.

Exercise

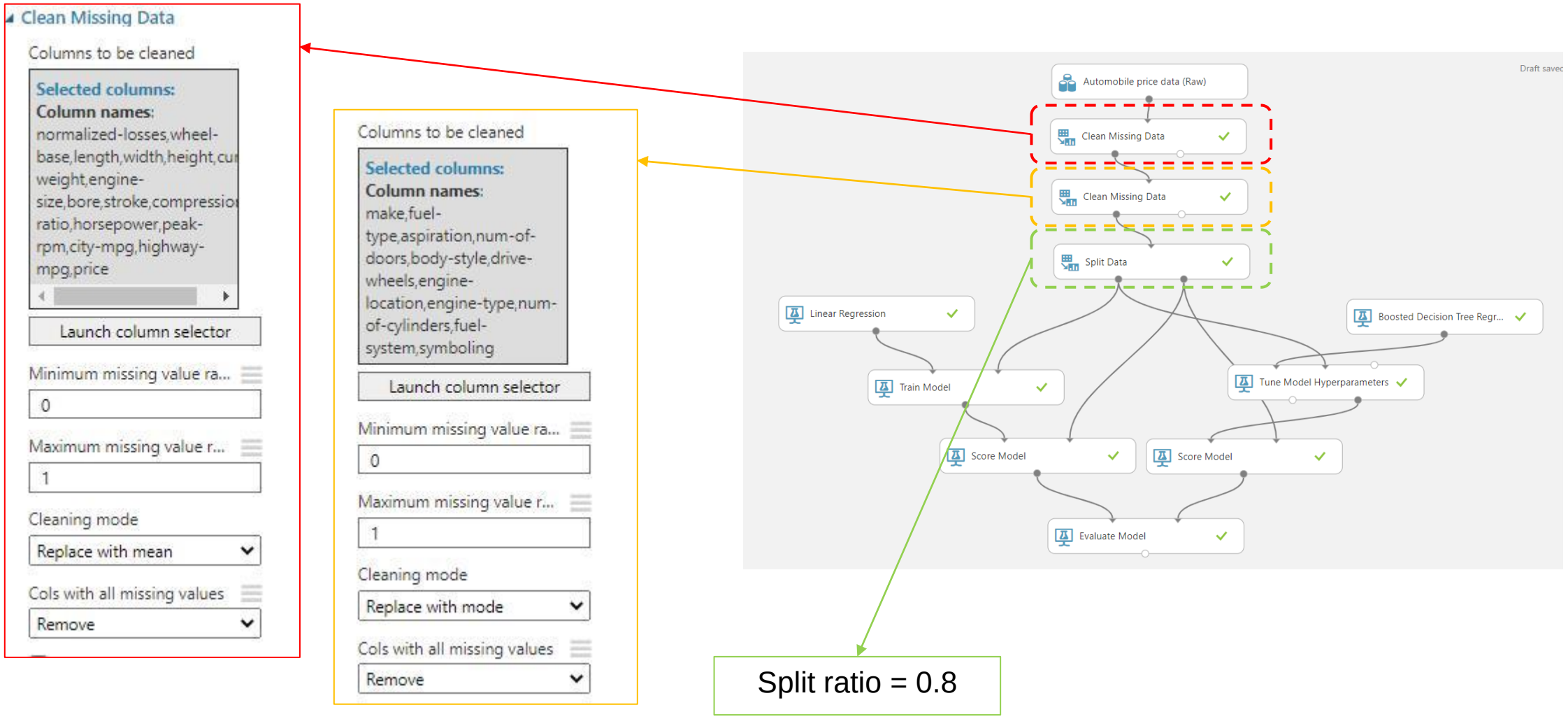
PETRONAS

Recreate linear regression and boosted decision tree flow

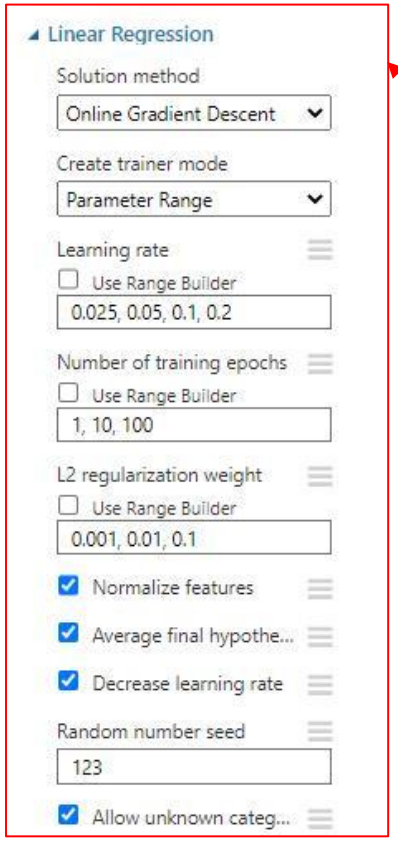


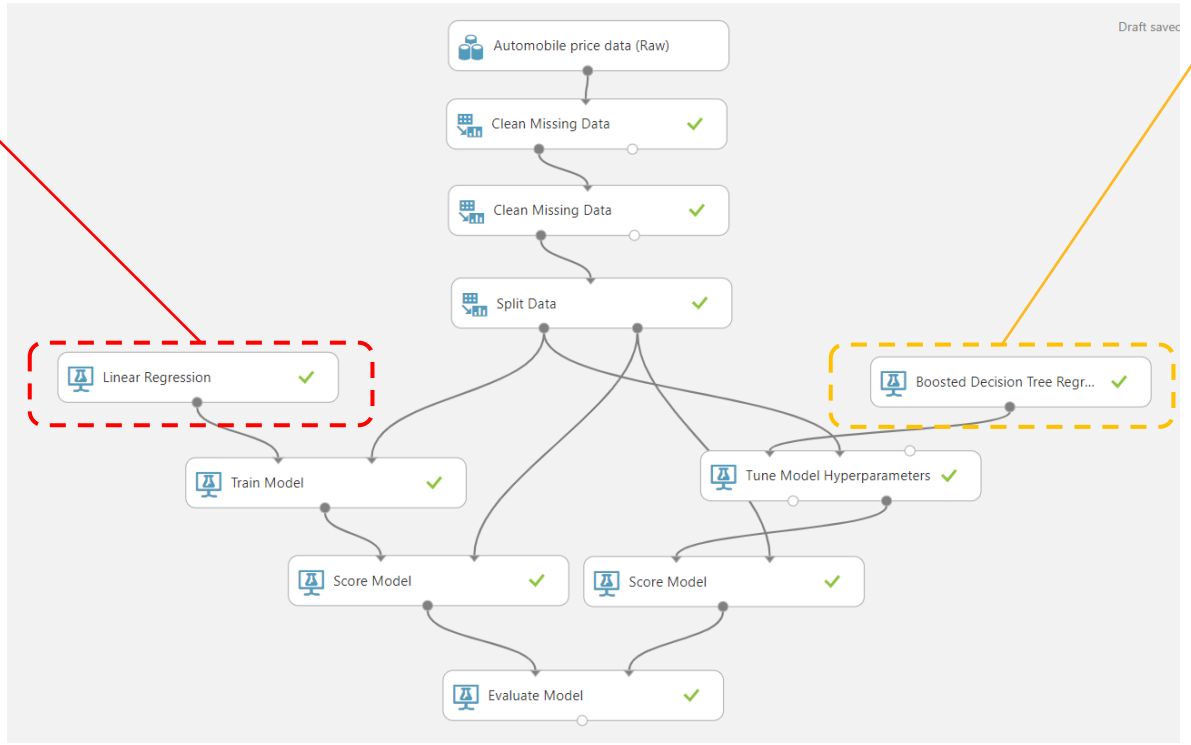
PETRONAS

Clean missing data details

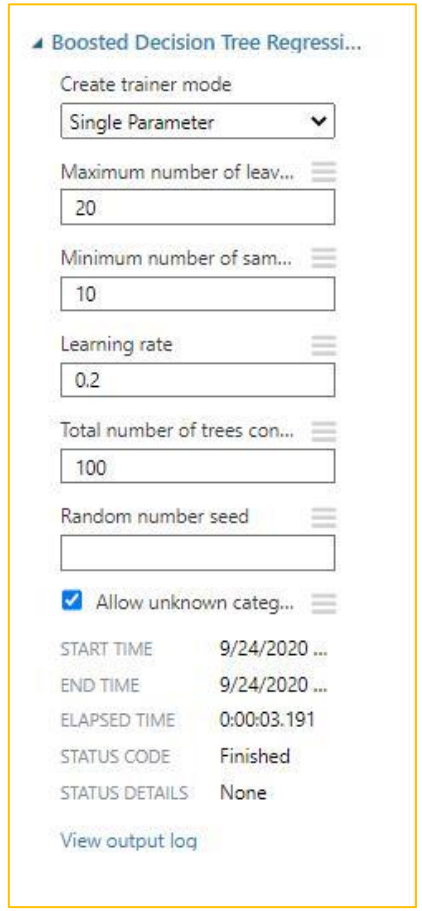


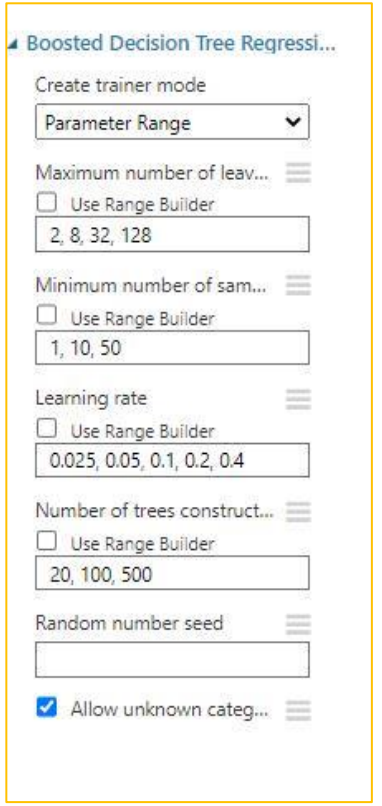
Regression details – Target variable for training is PRICE





Try both





Target variable for training is **PRICE**

References

References

Regression in Azure ML	https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/machine-learning-initialize-model-regression
Linear Regression	https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/linear-regression
Boosted Decision Tree Regression	https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/boosted-decision-tree-regression
Decision Forest Regression	https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/decision-forest-regression

Thank you for your passion!