

# Knowledge check

200 XP

2 minutes

1. You are creating a training pipeline for a regression model, using a dataset that has multiple numeric columns in which the values are on different scales. You want to transform the numeric columns so that the values are all on a similar scale based relative to the minimum and maximum values in each column. Which module should you add to the pipeline?

Select Columns in a Dataset

Normalize Data

**That is correct. When you need to transform numeric data to be on a similar scale, use a Normalize Data module.**

Clean Missing Data

2. Why do we split our data into training and validation sets?

Data is split into two sets in order to create two models, one model using the training set and a different model using the validation set.

Splitting data into two sets enables you to compare the labels that the model predicts with the actual known labels in the original dataset.

**That is correct. We want to test the model created with training data on validation data to see how well the model performs with data it was not trained on.**

We only need to split our data when we use the Azure Machine Learning Designer, not in other machine learning scenarios.

---

## Next unit: Summary