Question: evaluating Performance

What is cross-validation?

Cross-validation is a method of splitting all your data into two parts: training and validation. The training data is used to build the machine learning model, whereas the validation data is used to validate that the model is doing what is expected. This increases our ability to find and determine the underlying errors in a model.

* Training is special to the machine learning world. Because machine learning algorithms aim to map previous observations to outcomes, training is essential. These algorithms learn from data that has been collected, so without an initial set to train on the algorithm would be useless.

Swapping training with validation helps increase the number of tests. You would do this by splitting the data into two; the first time you'd use set 1 to train and set 2 to validate, and then you'd swap them for the second test. Depending on how much data you have you can split into more k-folds.

k-fold = 4 = 4 tests on four different partitions.