**Model Evaluation**

* **Cross Validation**
  + K-fold Cross Validation
    - Trys different sets of test and train data
    - Makes no sence, because we have historical data
* **Classification Accuracy**
  + We already use that
* **Logarithmic Loss**
  + We already have it, but never looked at it.
  + Best is 0
* **F-Measure**
  + considers both the precision and the recall
  + Precision is the number of correct positive results divided by the total predicted positive observations
  + Recall, on the other hand, is the number of correct positive results divided by the number of all relevant samples (total actual positives)
  + Makes no sense for our case?
* **Regression Metrics**
  + Mean Absolute Error (or MAE) is the sum of the absolute differences between predictions and actual values
  + Root Mean Squared Error (RMSE) measures the average magnitude of the error by taking the square root of the average of squared differences between prediction and actual observation
  + Makes sense?

Evaluate Model:

Accuracy, Precision, Recall, F1 Score, AUC (area under curve, higher is better), Threshold, true positive, true negative, false positive, false negative