Part 2 introduces the idea of measuring the absolute and incremental financial value of new information.

Absolute Financial Value of New Data

After determining that the Eggertopia scores have a high sustainable AUC, they should be evaluated exactly the same way as you used to evaluate your own model.

Find the threshold that minimizes cost-per-event on the Training Set, then use that threshold on the Test Set to identify the sustainable cost per event.

Eggertopia scores have an **AUC of .85** on both the training and test sets.

At the optimum threshold of .1 on the **Training Set**, the cost-per-event relying upon Eggertopia scores would be $600 [Cell BX17].

Making a new copy of the AUC Calculator for Test Set scores and outcomes, and applying the old Training Set threshold of .1 to the **Test Set** gives a new "sustainable" cost-per-event of **$838** [Cell BX17] (actually $837.50 exactly).

The financial value of Eggertopia scores, when compared against no model at all, is $1,250 (the base rate of default multiplied by the cost per false negative or .25%*$5000) minus the cost per event when using Eggertopia scores and the threshold determined on the Training Set.

Savings per event would be ($1,250 - $838) = **$412** (or $412.50)

Positive Predictive Value

It is useful to be able to calculate, for any predictive model, the probability of a default, given that the model classification is positive. This conditional probability is called the "Positive Predictive Value" or PPV.

Using the letter designations found on the **Information Gain Calculator** Spreadsheet: PPV = e/c. e is equal to the true positives, and c is the "test incidence" which is equal to the sum of the true positives plus the false positives.

The number of true positives is equal to the product of the true positive rate [from the AUC Calculator Spreadsheet, cell BX12] and the condition incidence of 25%.

The number of false positives is equal to the product of the false positive rate [From the AUC Calculator Spreadsheet, Cell BX14] and the no-condition incidence of 75%.

(.72*.25)/(.72*.25) + (.26*.75) = .48.

What this means is that with no model, any applicant has a 25% chance of defaulting – but the Eggertopia scores can be used to identify a group of applicants who have a 48% chance of defaulting, and weed them out.

Incremental Financial Value of New Data Versus Your Model

Part 2 of the Final Project concludes with you calculating the true *incremental* financial value of Eggertopia scores when compared to your own model.

For example, the forecasting model used to illustrate the sample answer in Part 1 had a cost-per-event of **$875**. This suggests that with an effective model in place, the incremental value of Eggertopia scores may be low.

In this example, it is only ($875 - $838) = **$37 per event.**

The idea to take away is that when the sales representatives from Eggertopia try to sell their scores, they will claim they are worth about $400 per event – but a little work by you can demonstrate that their true value to the bank is less than $40 per event.