Confusion Matrix and Metrics

When running a classification model, our resulting outcome is usually a binary 0 or 1 result, with 0 meaning False and 1 meaning True. We can compare our resulting classification outcomes with our actual values of the given observation to judge the performance of the classification model. The matrix used to reflect these outcomes is known as a **Confusion Matrix**, and can be seen below:

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Confusion matrices can be used to calculate performance metrics for classification models. Of the many performance metrics used, the most common are accuracy, precision, recall, and F1 score.

Accuracy:
The formula for calculating accuracy, based on the chart above, is ***(TP+TN)/(TP+FP+FN+TN)*** or all true positive and true negative cases divided by the number of all cases.

Accuracy is commonly used to judge model performance, however, there are a few drawbacks that must be considered before using accuracy liberally. One of these drawbacks deals with unbalanced datasets where one class, either true or false, is more common than the other causing the model to classify observations based on this imbalance. For example, if 90% of cases are false and only 10% are true, there's a very high possibility of our model having an accuracy score of around 90%. Naively, it may seem like we a have high rate of accuracy, but in actuality, we are just 90% likely to predict the 'false' class, so we don't actually have a good metric. Normally, I wouldn't use accuracy as a performance metric, I'd rather use precision, recall, or the F1 score.

Precision:
Precision is the measure of true positives over the number of total positives predicted by your model. The formula for precision can be written as: ***TP/(TP+FP).*** What this metric allows you to calculate is the rate of which your positive predictions are actually positive.

Recall:
Recall (a.k.a sensitivity) is the measure of your true positive over the count of actual positive outcomes. The formula for recall can be expressed as: ***TP/(TP+FN).*** Using this formula, we can assess how well our model is able to identify the actual true result.

F1Score:
The F1 score is the harmonic mean between precision and recall. The formula for the F1 score can be expressed as: ***2(p*r)/(p+r)*** where 'p' is precision and 'r' is recall. This score can be used as an overall metric that incorporates both precision and recall. The reason we use the harmonic mean as opposed to the regular mean, is that the harmonic mean punishes values that are further apart.