1. Evaluation Metrics

Precision

Precision is the ratio of correctly predicted positive observations to the total predicted positives. It tells us how many of the predicted positive instances are actually positive.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall (Sensitivity)

Recall is the ratio of correctly predicted positive observations to the all observations in actual class. It shows how many of the actual positive instances were correctly identified by the model.

$$\text{Recall} = \frac{TP}{TP+FN}$$

Accuracy

Accuracy is the ratio of correctly predicted observations to the total observations. It's a useful metric when the classes are balanced.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

2. Confusion Matrix

A confusion matrix is a table used to evaluate the performance of a classification model. It provides the counts of true positive, true negative, false positive, and false negative predictions.

Example

A confusion matrix for a binary classification problem:

	Predicted Positive	**Predicted Negative**
Actual Positive	TP	FN
Actual Negative	FP	TN

3. Heatmaps

Heatmaps are a great way to visualize the confusion matrix and understand the performance of your model.

4. F1 Score

The F1 Score is the harmonic mean of precision and recall, providing a single metric that balances both concerns.

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. ROC Curve and AUC

The ROC (Receiver Operating Characteristic) curve is a graphical representation of the true positive rate (recall) versus the false positive rate. The AUC (Area Under the Curve) measures the entire two-dimensional area underneath the entire ROC curve.

6. Cross-Validation

Cross-validation involves dividing the dataset into k subsets and training the model k times, each time using a different subset as the validation set and the remaining as the training set. This helps in obtaining a more robust estimate of the model's performance.