

# Confusion Matrix

	1 (Predicted)	0 (Predicted)
1 (Actual)	True Positive	False Negative
0 (Actual)	False Positive	True Negative

**Confusion matrix** is used to measure the performance of the logistic regression. It has two rows and two columns.

- The two rows represent the actual events; the first row when the event happened, and the second row when the event didn't happen.
- The two columns represent the prediction of the logistic regression; the first column when the logistic regression predicts the event will happen and the second column when the logistic regression predict that event will not happen
- There are intersections between the rows and columns;
  - the first intersection is between the actual one and the predicted one, referred to as **True Positive**. **True** means that the predictions come true. **Positive** means it happened.
  - the second intersection is between the actual one and the predicted zero, referred to as **False Negative**. **False** means that the prediction is wrong or false and **Negative** means logistic regression predicted something not to happen and actually it happened.
  - the third intersection is between actual zero and predicted one, referred to as **False Positive**. **False** means the prediction is wrong, and predicted something to happen and it didn't happen. **Positive** means a prediction of something to happen and actually didn't happen so that's why the prediction is positive but actually it's negative or it didn't happen.
  - At the fourth intersection, both the prediction and actual are zero, referred to as **True Negative**. **True** means the prediction comes true. **Negative** means the event did not happen.

## Example

### 100 Customers

		Predicted		
		0	1	
Actual	0	52	5	57
	1	7	36	43
		59	41	100

- First row =  $52 + 5 = 57$
- Second row =  $7 + 36 = 43$ 
  - The total customer is 100 ( $57 + 43$ )
- The predicted values =  $52 + 7 = 59$ , and  $5 + 36 = 41$ 
  - The total customer is 100 ( $59 + 41$ )

### 100 Customers

		Predicted		
		0	1	
Actual	0	52	5	57
	1	7	36	43
		59	41	100

- Logistic regression predicted 52 and 36 as the only true and correct values.
- Logistic regression predicted that the customer will not buy and they did not buy and 36 customers will buy and they bought.

$$52 + 36 = 88$$

$$88/100 = 88\%$$

The accuracy of the logistic regression is 88%, hence a good model

- The model above 80% is good
- The model above 90% is excellent

