Exercise 1:

(a) Below is an empty confusion matrix for a binary classification problem. Assign the correct terms (TP, FP, FN, TN) to fields (A), (B), (C), (D) and explain what each term represents.

	Actual Positive	**Actual Negative**
Predicted Positive	(A)	(B)
Predicted Negative	(C)	(D)

(b) Below is a table that incorrectly assigns 11 metrics derived from the confusion matrix to their corresponding formulas and descriptions. The table contains the following columns:

- Metric Names (Lettered A-O): Commonly used names (including alternative names) for each metric.
- Formula (Lettered a-o): Expressions using TP, FP, FN, TN.
- Description (Numbered 1-11): Brief descriptions of metrics derived from the confusion matrix.

Your task is to:

- Correctly match metric names, formulas, and descriptions in a new table.
- Identify metric names that are synonyms (i.e., referring to the same metric).
- Identify any formula that does not correspond to a metric or description and list them separately.

Metric Name	**Formula**	**Description**
A) True Positive Rate (TPR)	a) $\frac{\text{FN}}{\text{TP+FN}}$	1) Proportion of actual positives correctly identified.
B) Recall	b) $\frac{\text{TP}}{\text{TP+FN}}$	2) Proportion of actual negatives correctly identified.
C) Sensitivity	c) $\frac{\text{TN}}{\text{TN+FP}}$	3) Proportion of positive predictions that are correct.
D) True Negative Rate (TNR)	d) $\frac{\text{TP}}{\text{TP+FP}}$	4) Proportion of positive predictions that are incorrect.
E) Specificity	e) $\frac{\text{FP}}{\text{TP+FP}}$	5) Proportion of actual negatives incorrectly classified as positive.
F) Precision	f) $\frac{\text{FP}}{\text{FP+TN}}$	6) Overall proportion of correct predictions.
G) Positive Predictive Value (PPV)	g) $\frac{\text{TP+TN}}{\text{TP+FP+FN+TN}}$	7) Combines precision and recall using their harmonic mean.
H) False Discovery Rate (FDR)	h) $\frac{2\cdot\text{Precision}\cdot\text{Recall}}{\text{Precision+Recall}}$	8) Proportion of actual positives in the dataset.
I) False Positive Rate (FPR)	i) $\frac{2\cdot(\text{Precision+Recall})}{\text{Precision}\cdot\text{Recall}}$	9) Proportion of negative predictions that are correct.
J) False Negative Rate (FNR)	j) $\frac{\text{TP+FN}}{\text{TP+FP+FN+TN}}$	10) Proportion of negative predictions that are incorrect.
K) Negative Predictive Value (NPV)	k) $\frac{\text{TN}}{\text{TN+FN}}$	11) Proportion of actual positives incorrectly classified as negative.
L) Accuracy	l) $\frac{\text{FN}}{\text{TN+FN}}$	
M) False Omission Rate (FOR)	m) $\frac{\text{FN}}{\text{FP+FN}}$	
N) Prevalence	n) $\frac{\text{FN+FP}}{\text{TP+TN+FP+FN}}$	
O) F1 Score	o) $\frac{\text{TP−FP}}{\text{TP+FN+FP+TN}}$	

Solution 1:

(a) **Confusion Matrix Completion and Explanation**

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

- TP: Cases where the model correctly predicts the positive class.
- FP: Cases where the model incorrectly predicts the positive class for a negative actual value.
- FN: Cases where the model incorrectly predicts the negative class for a positive actual value.
- TN: Cases where the model correctly predicts the negative class.

(b) Below a corrected table with properly assigned metric names, formulas, and descriptions:

Metric Name	**Formula**	**Description**
A) True Positive Rate (TPR) B) Recall C) Sensitivity	b) $\frac{TP}{TP+FN}$	1) Proportion of actual positives correctly identified.
D) True Negative Rate (TNR) E) Specificity	c) $\frac{TN}{TN+FP}$	2) Proportion of actual negatives correctly identified.
F) Precision G) Positive Predictive Value (PPV)	d) $\frac{TP}{TP+FP}$	3) Proportion of positive predictions that are correct.
H) False Discovery Rate (FDR)	e) $\frac{FP}{TP+FP}$	4) Proportion of positive predictions that are incorrect.
I) False Positive Rate (FPR)	f) $\frac{FP}{FP+TN}$	5) Proportion of actual negatives incorrectly classified as positive.
J) False Negative Rate (FNR)	a) $\frac{FN}{TP+FN}$	11) Proportion of actual positives incorrectly classified as negative.
K) Negative Predictive Value (NPV)	k) $\frac{TN}{TN+FN}$	9) Proportion of negative predictions that are correct.
L) Accuracy	g) $\frac{TP+TN}{TP+FP+FN+TN}$	6) Overall proportion of correct predictions.
M) False Omission Rate (FOR)	l) $\frac{FN}{FN+TN}$	10) Proportion of negative predictions that are incorrect.
N) Prevalence	j) $\frac{TP+FN}{TP+FP+FN+TN}$	8) Proportion of actual positives in the dataset.
O) F1 Score	h) $\frac{2 \cdot Precision \cdot Recall}{Precision+Recall}$	7) Combines precision and recall using their harmonic mean.

Identified Synonyms:

- True Positive Rate (TPR), Recall, and Sensitivity refer to the same metric.
- True Negative Rate (TNR) and Specificity refer to the same metric.
- Precision and Positive Predictive Value (PPV) refer to the same metric.

Unassignable Formulas:

- $i) \frac{2 \cdot (Precision+Recall)}{Precision \cdot Recall}$
- $o) \frac{TP-FP}{TP+FN+FP+TN}$
- $n) \frac{FN+FP}{TP+TN+FP+FN}$
- $m) \frac{FN}{FP+FN}$