Building Confusion Matrix

As we are planning to create a confusion matrix, so firstly we have to compare the predicted diabetes-2 to the actual diabetes-2. Based on LR outcome, our recall score of 0.761 and precision of 0.633, we can use these evaluation metrics to build the confusion matrix for justifying our model further in CBA.

So let's look at the terms how do we define it-

True Positives **(TP):** Correctly predicted diabetes 2 cases.
False Positives **(FP):** Incorrectly predicted diabetes 2 cases (predicted as diabetes but not actually diabetes).
True Negatives **(TN):** Correctly predicted non-diabetes 0 and pre-diabetes 1 cases.
False Negatives **(FN):** Incorrectly predicted non-diabetes 0 & pre-diabetes 1 cases (predicted as non-diabetic & pre-diabetes cases but actually diabetes 2).

So from our data, we can build the following equation:

Total actual diabetes 2 cases, TP+FN = 35346 .....(1)
Total dataset size, (TP+FP+TN+FN) = 253680 .....(2)
Recall, TP / (TP + FN) = 0.761 .....(3)
Precision, TP / (TP + FP) = 0.633 .....(4)

From equation 1 & 3, we can calculate the TP = 35346*0.761
= 26898

Using precision, we can calculate FP (From equation 4).

Precision = TP / (TP + FP)
0.633 = 26898/ (26898 + FP)
FP = 26898/0.633 − 26898
FP = 15595

Now, from equation 1, we can calculate the FN = 35346-26898
= 8448

And now we can calculate TN (From equation 2).

TP+FP+TN+FN = 253680
TN = 253680 − (TP+FP+FN)
TN = 253680 − (26898+15595+8448)
TN = 202739

Now put these values in the followings and create the confusion matrix-

Confusion matrix	**Predicted No Diabetes Or, Pre-diabetes-1**	**Predicted Diabetes-2**
Actual (No Diabetes Or, Pre Diabetes)	**202739**	**15595**
Actual Diabetes	**8448**	**26898**

This confusion matrix indicates:
True Negatives (TN): 202739
False Negatives (FN): 8448
False Positives (FP): 15595
True Positives (TP): 26898