EDA

1. Provider Dataset

Shape of th Provider Dataset: (5410, 2) Number of features in the Provider Dataset: 1

* There are a total of 5410 datapoints, each signifying a Provider ID and a Class Label 'PotentialFraud' indicating whether the given Provider is fraudulent or not.
* Based on the dataframe.info() method, we can see that there are **no missing values** in the Provider dataset.
* This is a case of **binary classification** task because the Class Label 'PotentialFraud' has only two unique values:

  i. No: Indicates that the Provider is non-fraudulent.  
  ii. Yes: Indicates that the Provider is fraudulent.

* The dataset is **highly imbalanced** with 90.6% non-fraud Providers and 9.4% fraud Providers class labels.