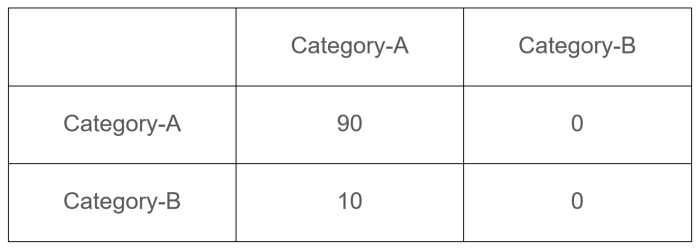
**What is an Imbalanced Dataset?**

Imagine, you have two categories in your dataset to predict — Category-A and Category-B. When Category-A is higher than Category-B or vice versa, you have a problem of imbalanced dataset.

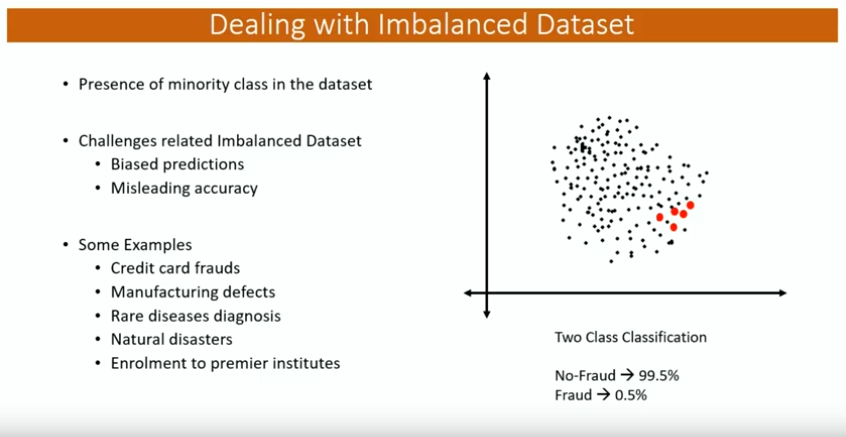
So how is this a problem?

Imagine in a dataset of 100 rows, Category-A is containing 90 records and Category-B is containing 10 records. You run a machine learning model and end up with 90% accuracy. You were excited until you checked the confusion matrix.



Confusion Matrix

Here, Category-B is completely classified as Category-A and the model got away with an accuracy of 90%.



Red dots denote minority class in the dataset, black dots denote majority class in the dataset.

