**Credit Risk Analysis Report:**

**Overview:**

Credit risk poses a classification problem that’s inherently imbalanced. The reason is that healthy loans easily outnumber risky loans. Various techniques to train and evaluate models with imbalanced classes has been examined. Coding involves dataset of historical lending activity from a peer-to-peer lending services company to build a model that can identify the creditworthiness of borrowers.

**Key takeaways**

* **Accuracy**(total % of correct classifications) is often not a meaningful performance metric when data sets are imbalanced.
* **Class imbalance** implies that one class is represented much more often than others. Penalizing certain misclassifications and re-sampling the data set help dealing with imbalances.
* **Precision**is a metric that penalizes false positives. As such, models with high precision are cautious to label an element as positive.
* **Recall**is a metric that penalizes false negatives. Models with high recall tend towards positive classification when in doubt.
* **F-scores** and **precision-recall curves** provide guidance into balancing precision and recall.
* In the exercise – For original data -Healthy loans are 100% precision while high risk loans is only 85%, same case with recall 0s are 99% versus just 91% recall for high risk loans.
* With the oversampled data, we can see an increase in recall, but the precision has dropped.