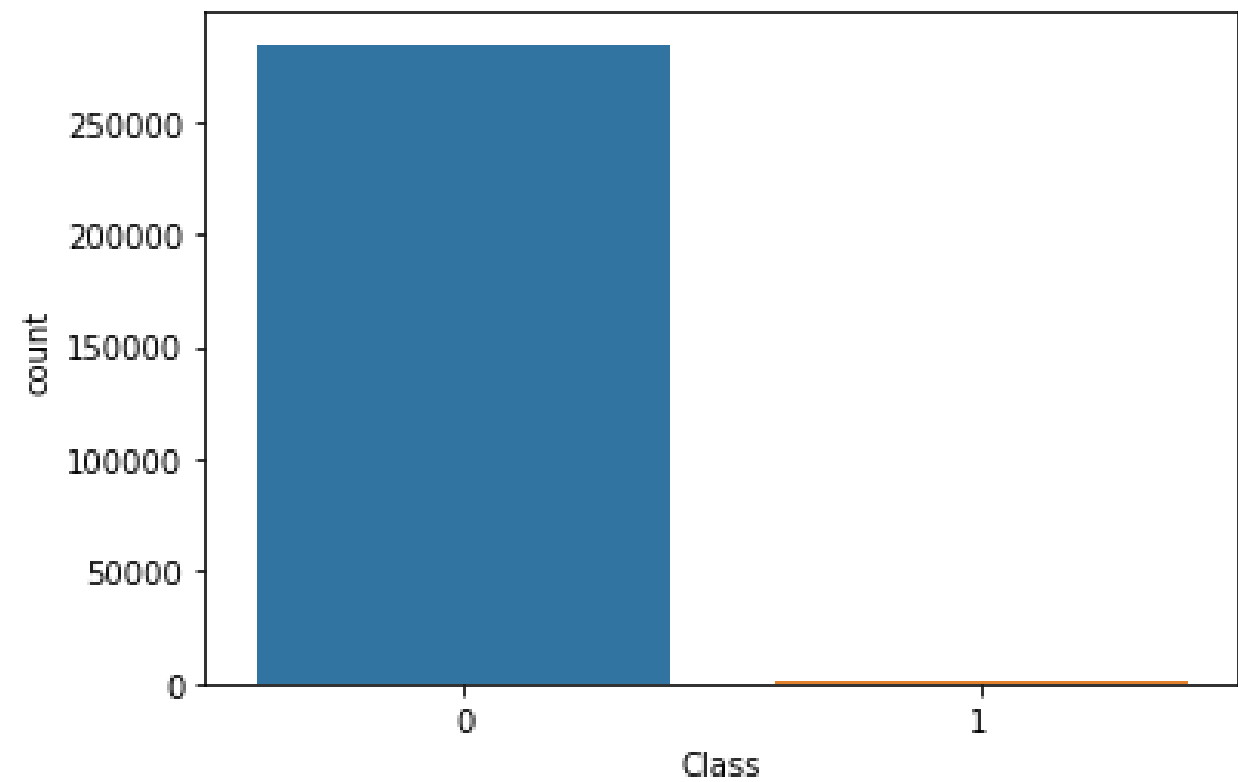


Exploring Class Imbalance with Fraud Detection

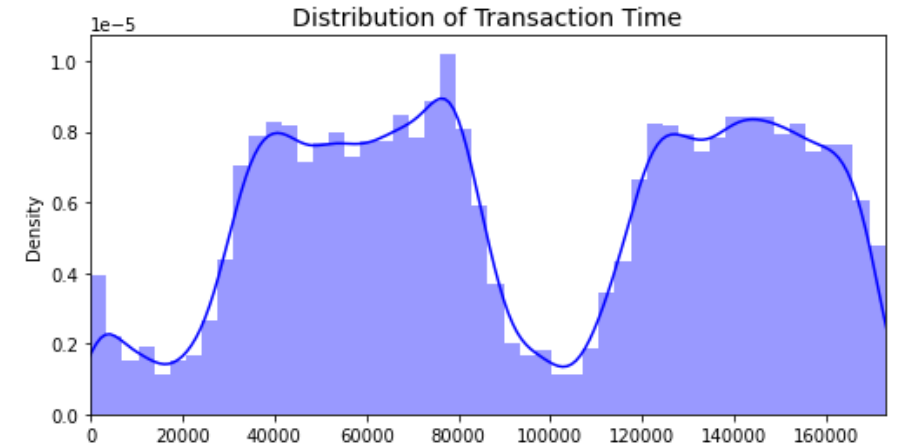
James Bush
Springboard
Milestone Report 1

Class Counts by
Dependent Variable

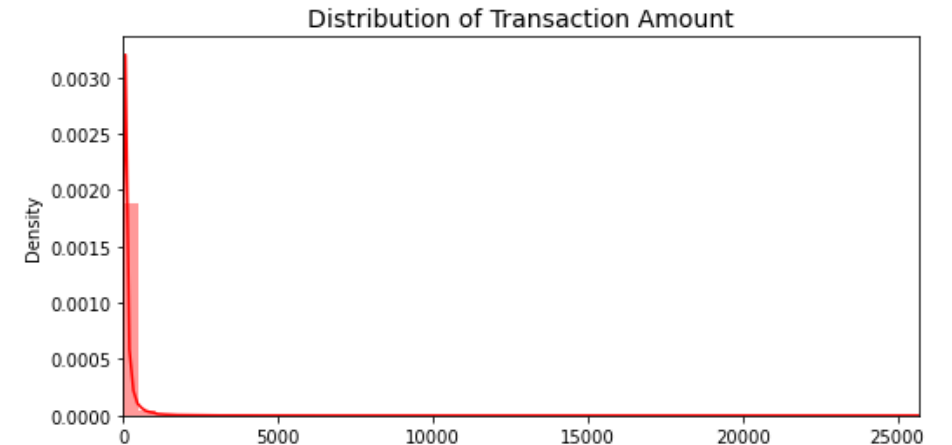


Untransformed Variables

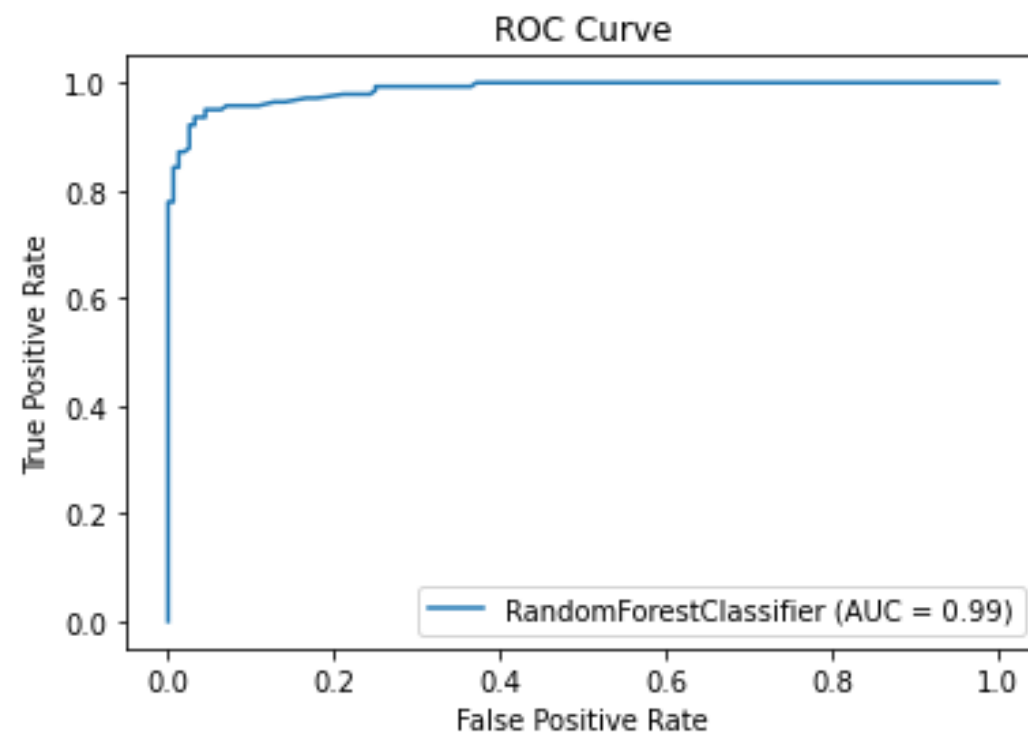
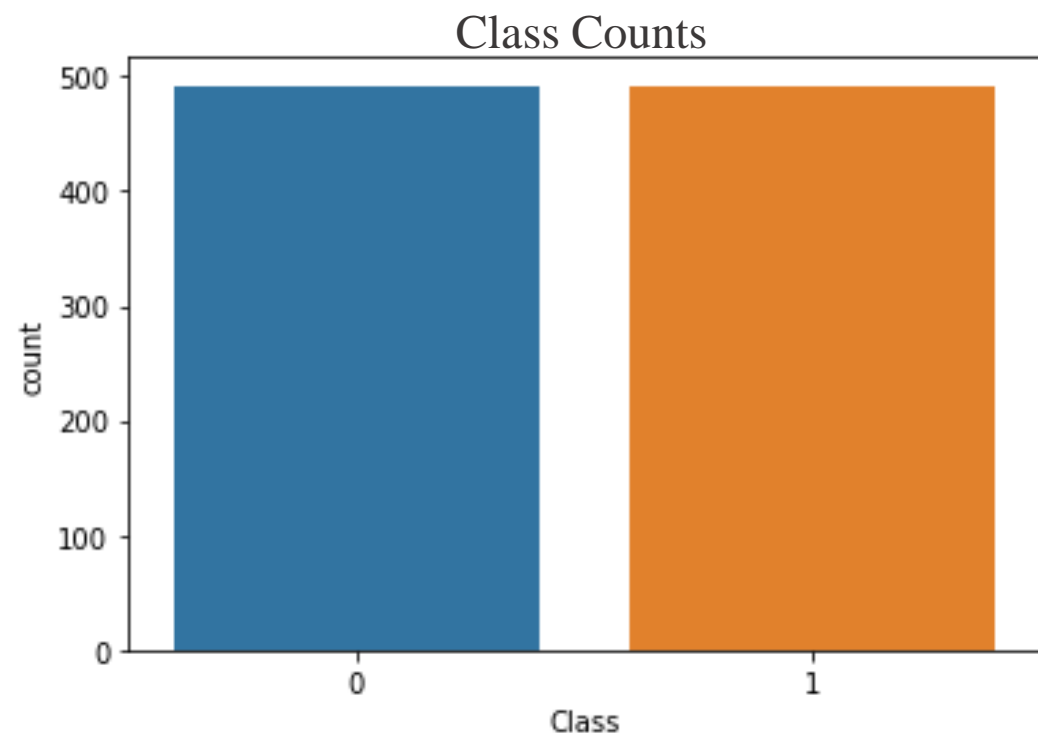
Time as the number of seconds elapsed between each transaction and the first.



Transaction Amount.
Note the one-tailed skew to the lower amount values.

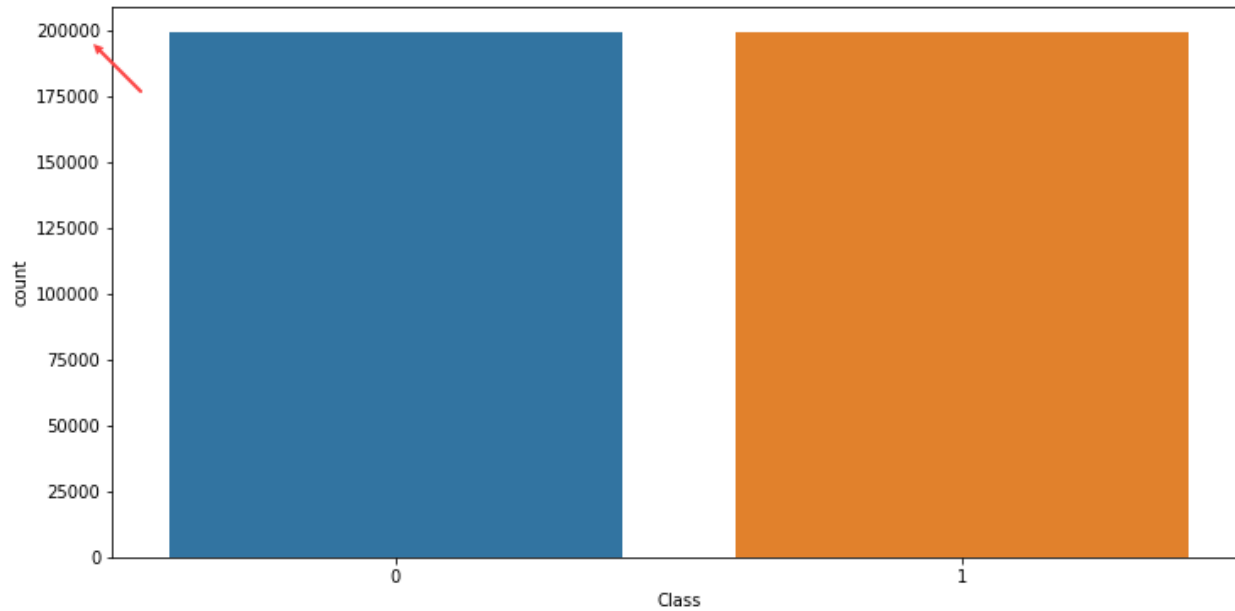


Undersampling

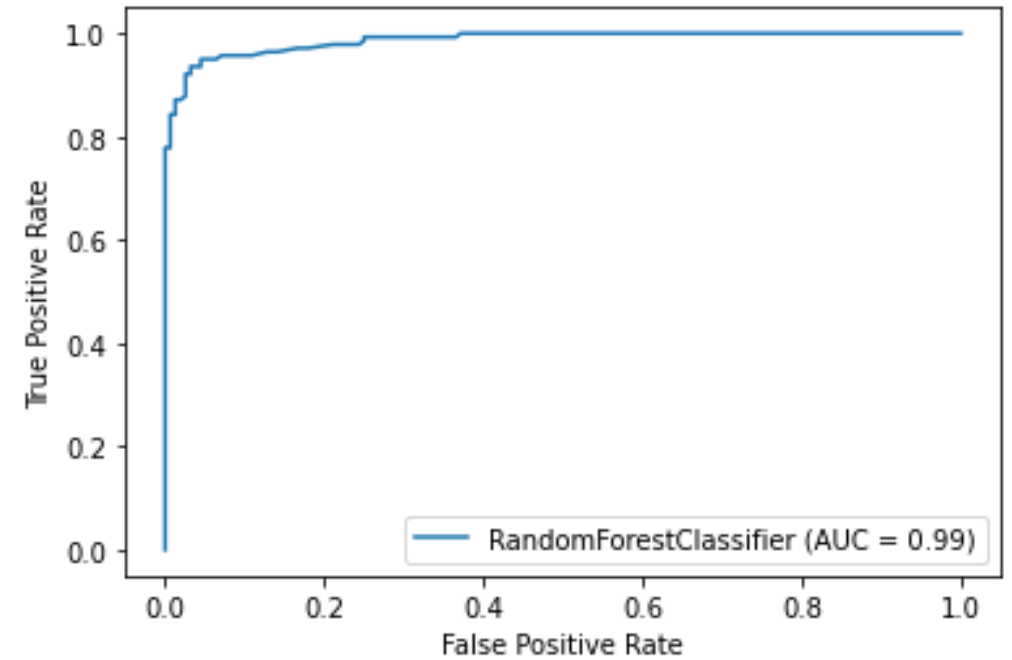


Oversampling

Class Counts

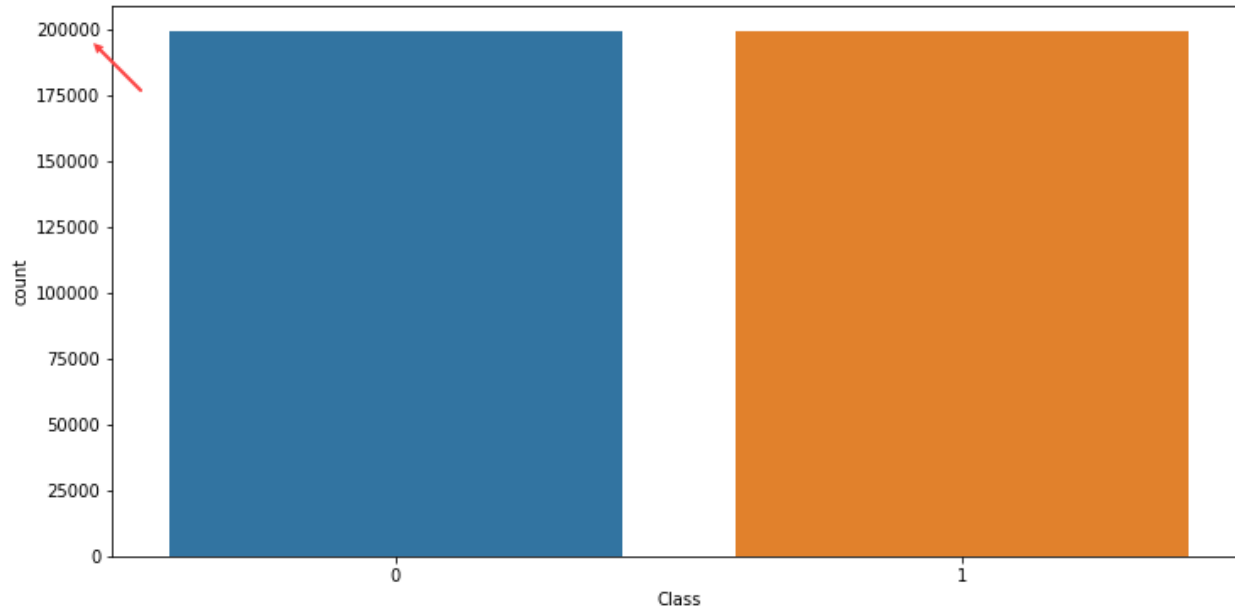


ROC Curve

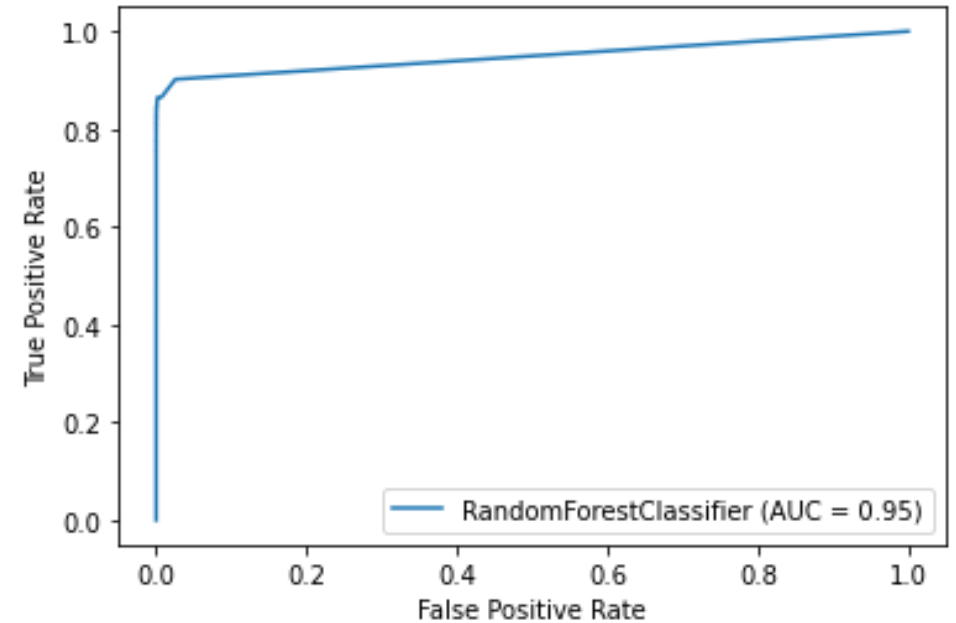


Synthetic Minority Oversampling Technique (SMOTE)

Class Counts



ROC Curve



Though the AUC is lower, SMOTE improves the Precision, Recall scores.

Model Evaluation

