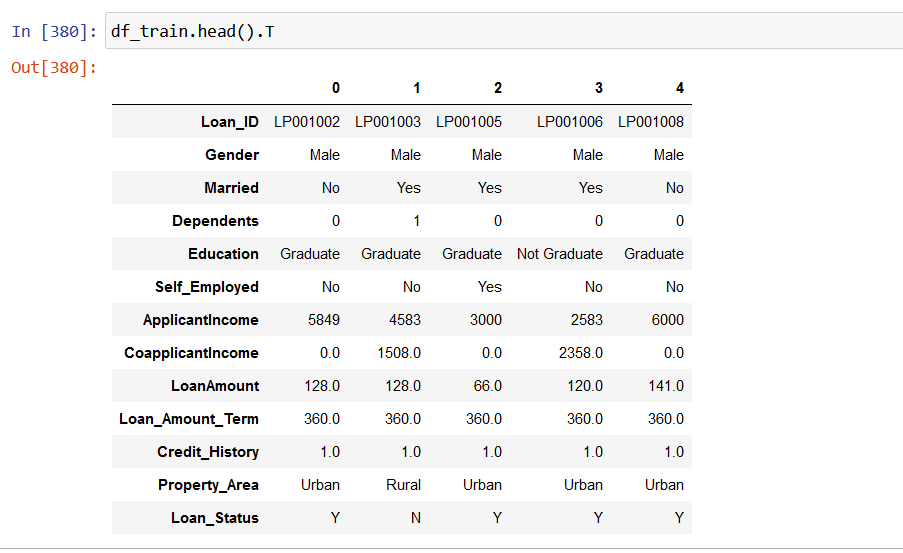
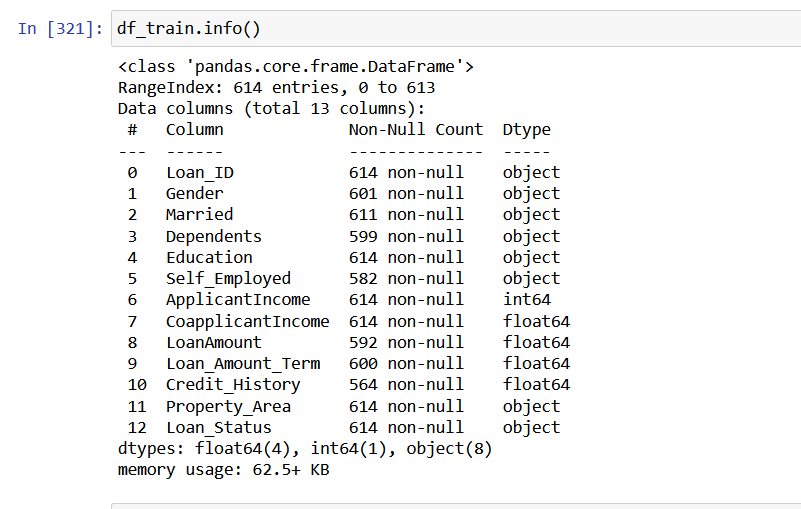
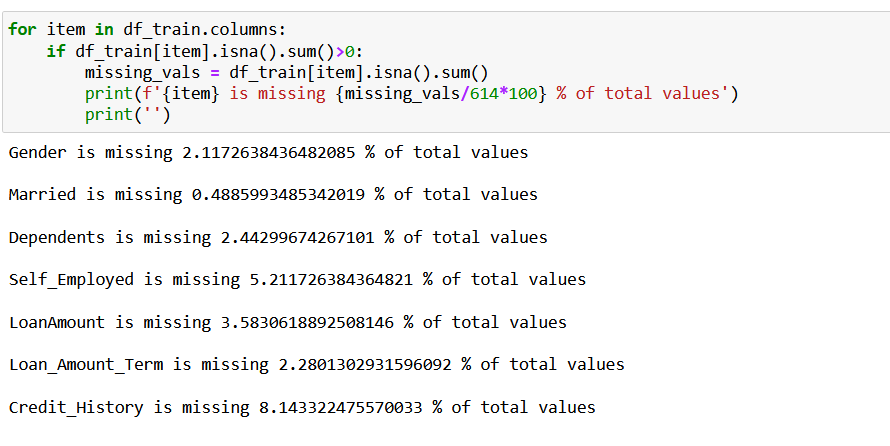
Dataset review:

<https://www.kaggle.com/code/ajinkyataranekr/loan-eligibility-prediction/data>



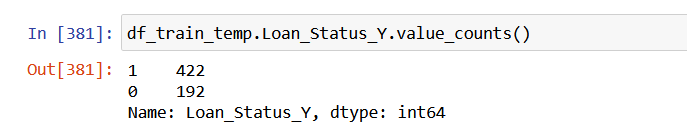


Data is quite small (614 observations) and has missing values which required imputation:

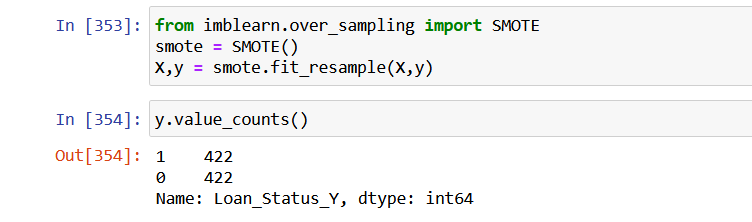


The target column was also unbalanced. SMOTE was used to balance the two classes:

Before SMOTE:



After SMOTE:



Model chosen:

DTree, RForest, Gboost and XGB were used initially to see which model performed the best. The top 2-3 models were used in the voting classifier

Model performance:

