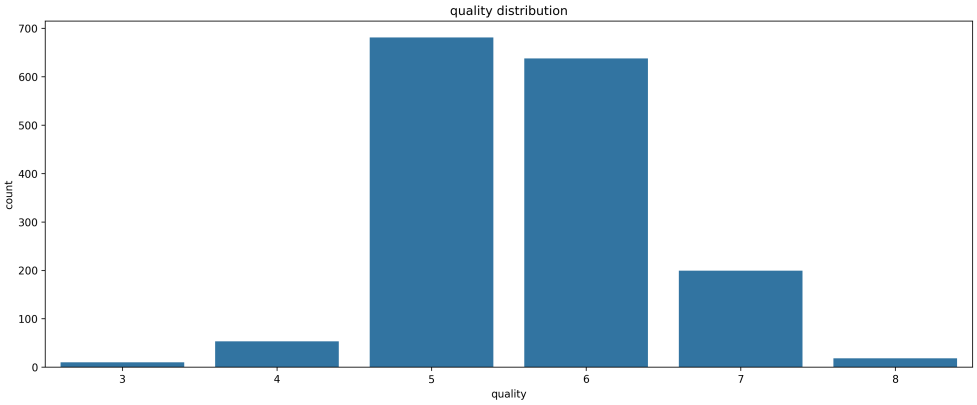
Wine Quality Automated Report

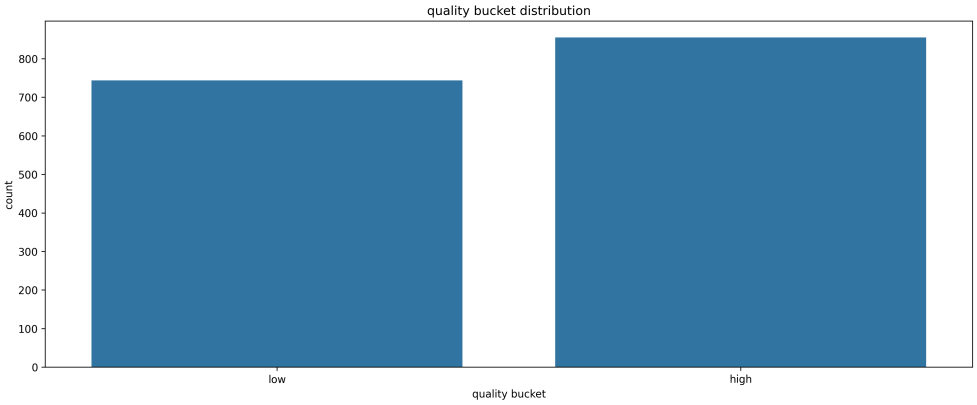
Author: Denitsa Panova
Date: 2023-11-13
Disclaimer: The objective of this report is to present the outcomes generated by four distinct machine learning learning techniques: correlation analysis, principal components analysis (PCA), random forest, and logistic regression. A specialized interpretation is essential to derive accurate conclusions regarding the chemical factors influencing the high-quality wine.

Dependant variable distribution

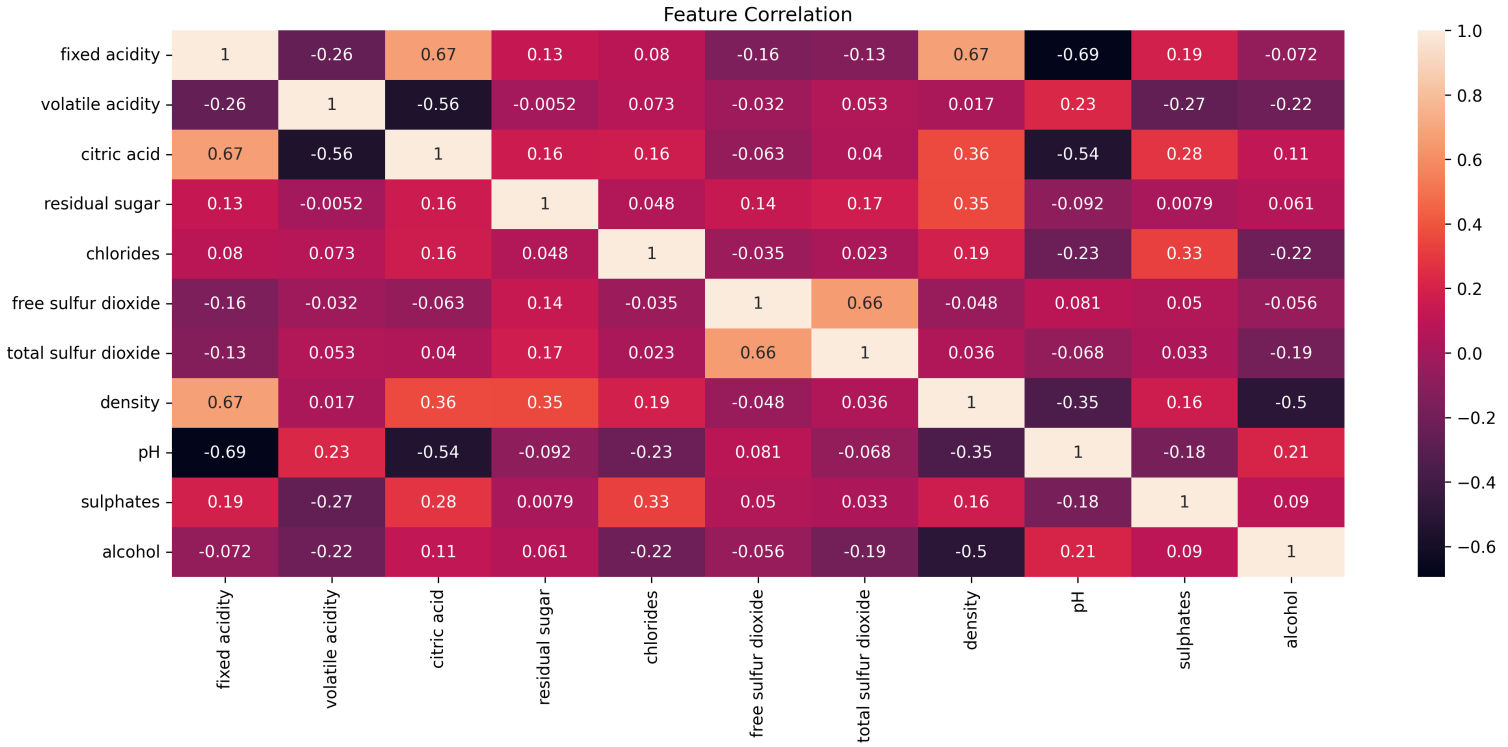
Original Variable



Created Variable



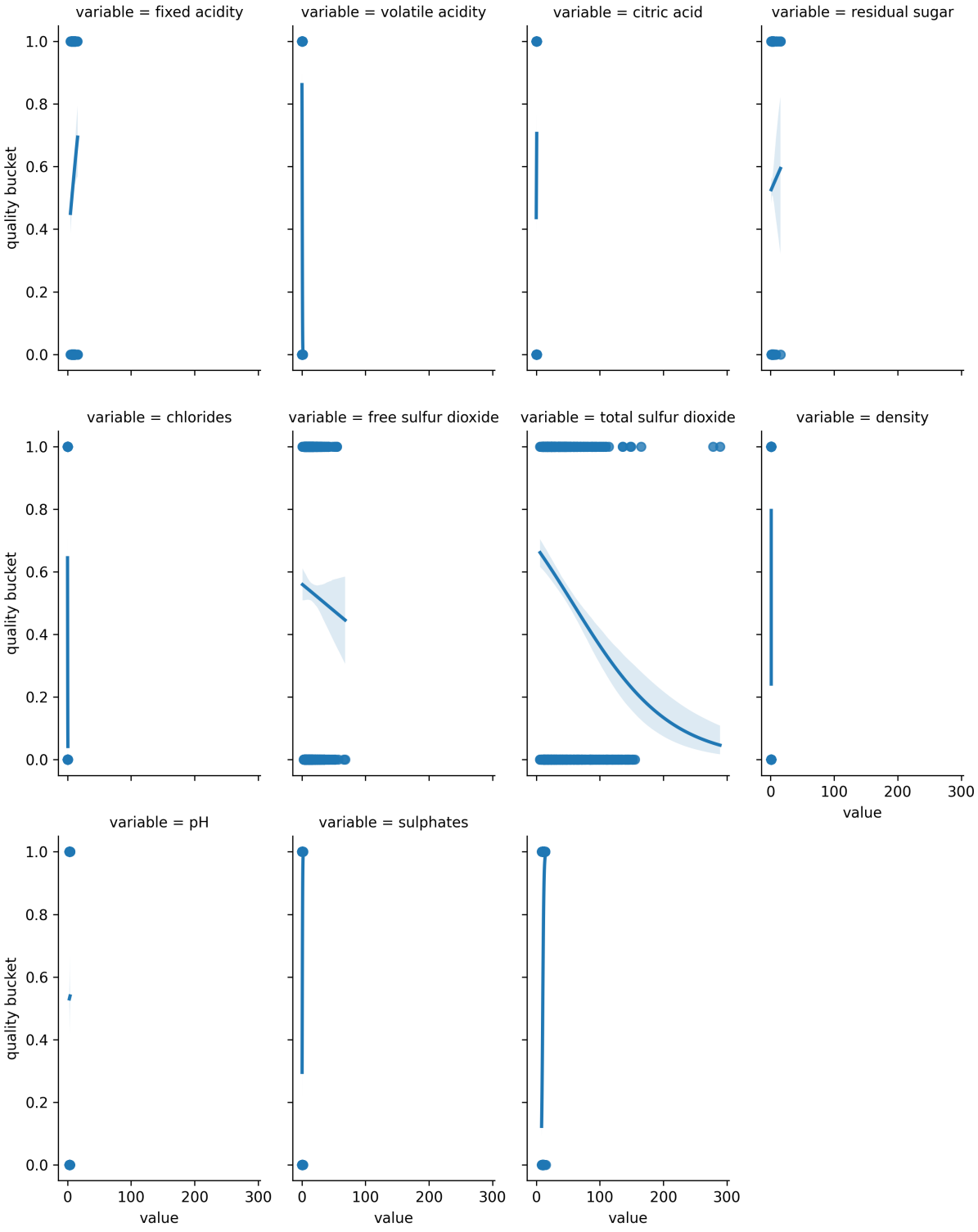
Correlation Analysis



PCA

The explained variance for the first principal component is 0.95

Independent Logistic Regressions



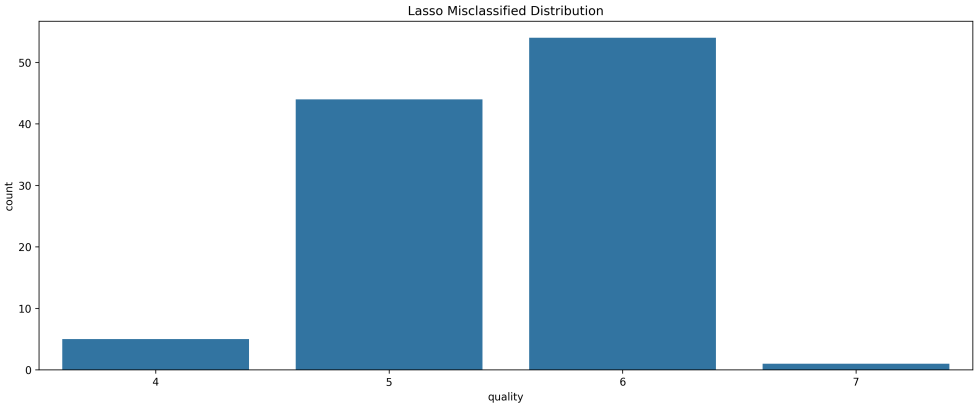
Lasso Logistic Regression Results

The model accuracy is 0.74

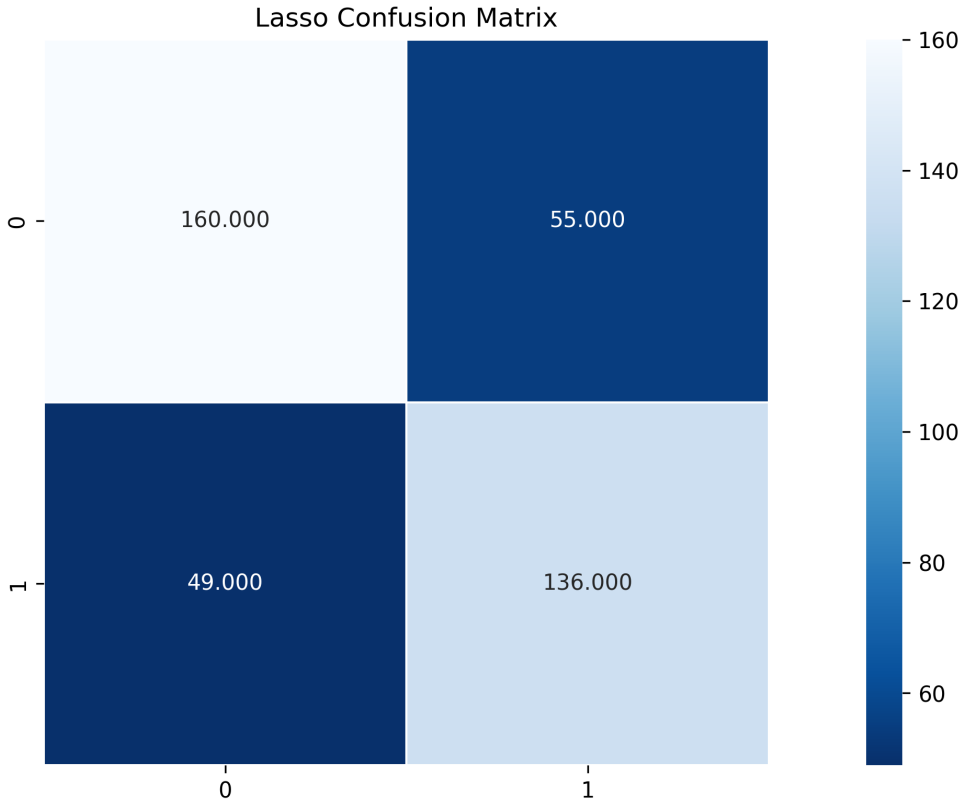
The model precision is 0.739

The model precision is 0.74

Below is a graph representing the distribution of the misclassified quality classes



Below is a graph representing the confusion matrix



Lasso Logistic Regression Drivers

Feature fixed acidity has a coefficient 0.103

Feature volatile acidity has a coefficient 2.524

Feature residual sugar has a coefficient -0.004

Feature free sulfur dioxide has a coefficient -0.028

Feature total sulfur dioxide has a coefficient 0.017

Feature pH has a coefficient 2.048

Feature sulphates has a coefficient -1.105

Feature alcohol has a coefficient -0.846

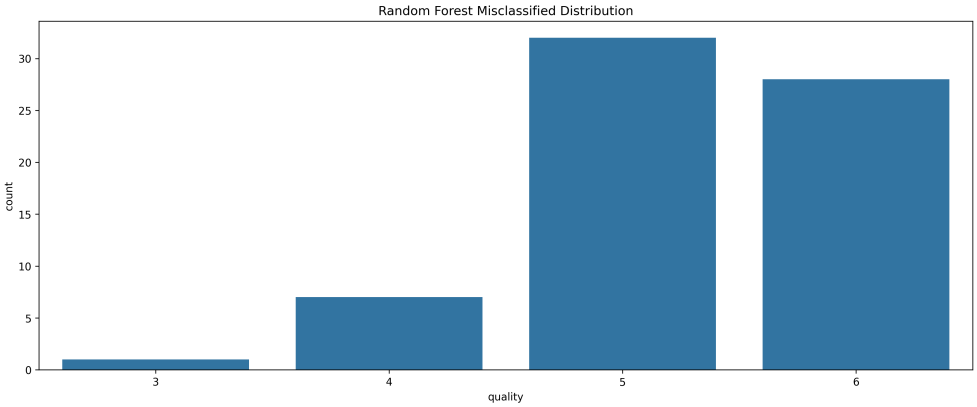
Random Forest Results

The model accuracy is 0.83

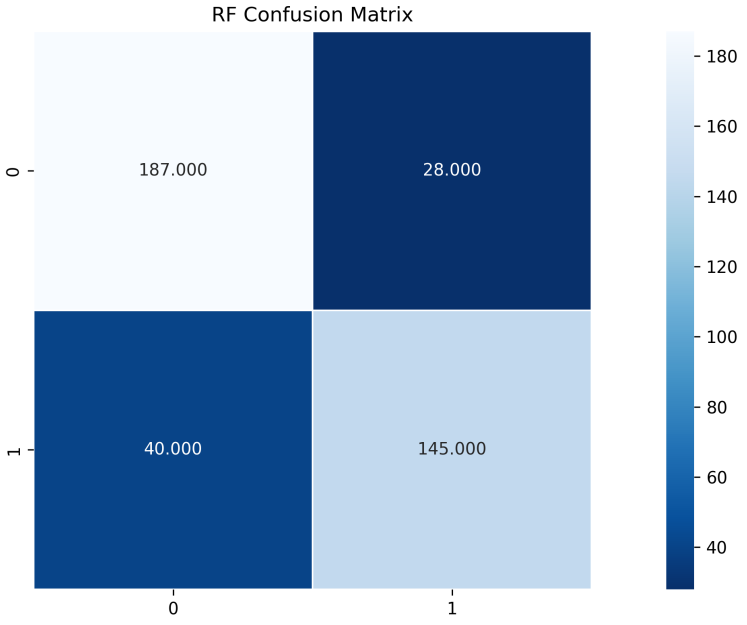
The model precision is 0.831

The model precision is 0.827

Below is a graph representing the distribution of the misclassified quality classes



Below is a graph representing the confusion matrix



Random Forest Drivers