

# Wine Quality Automated Report

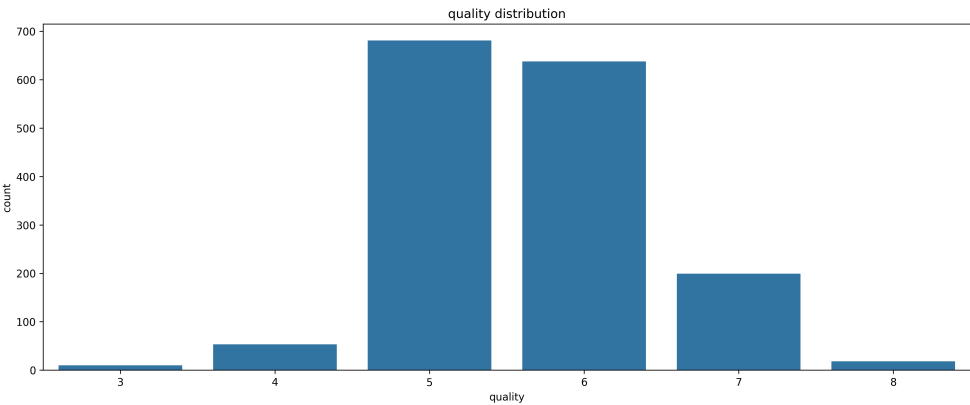
Author: Denitsa Panova

Date: 2023-11-13

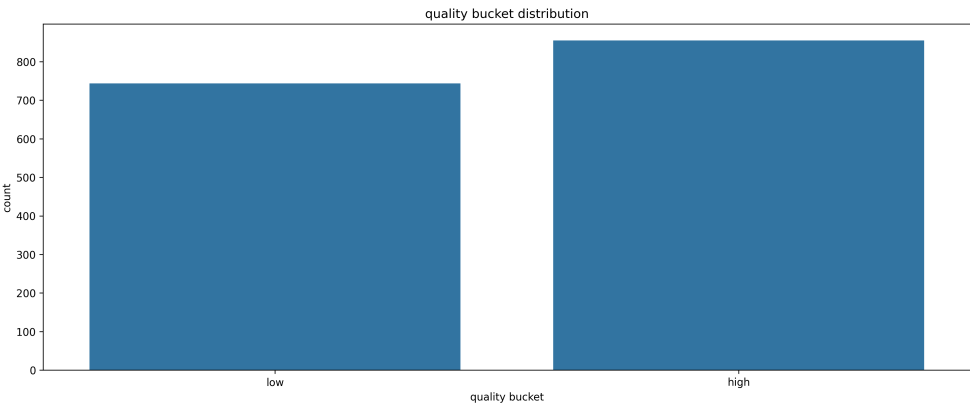
Disclaimer: The objective of this report is to present the outcomes generated by four distinct machine learning learning techniques: correlation analysis, principal components analysis (PCA), random forest, and logistic regression. A specialized interpretation is essential to derive accurate conclusions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

## Dependant variable distribution

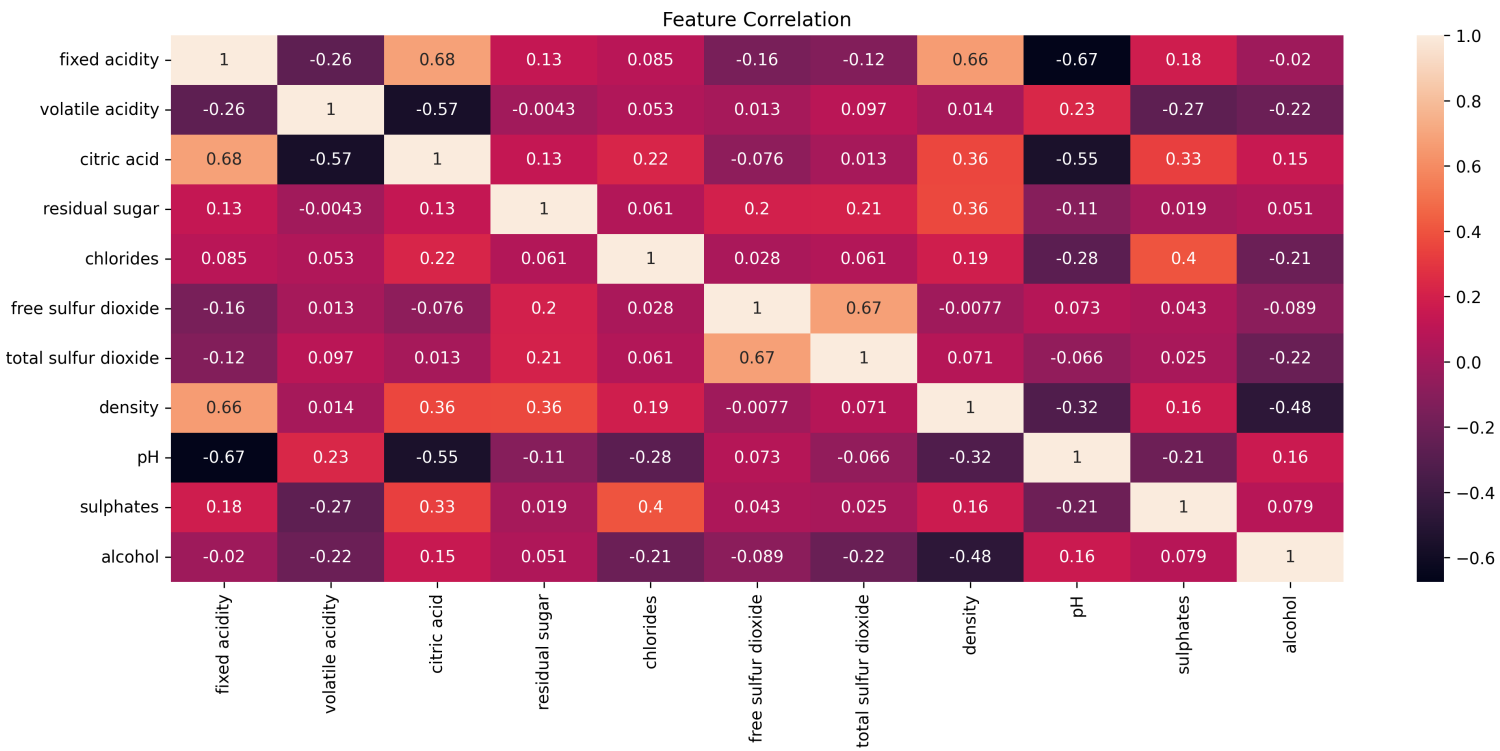
Original Variable



Created Variable



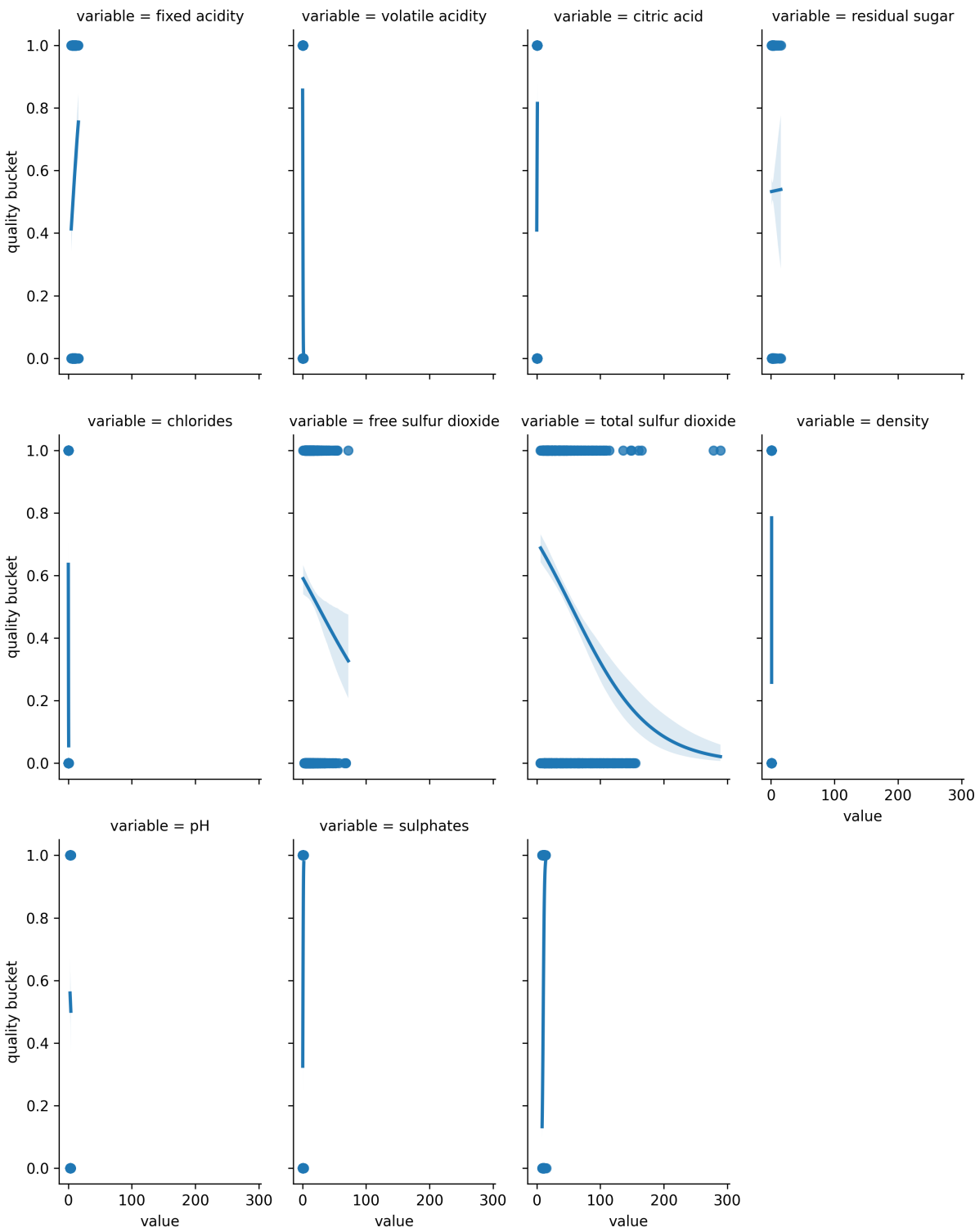
# Correlation Analysis



# PCA

The explained variance for the first principal component is 0.95

# Independent Logistic Regressions



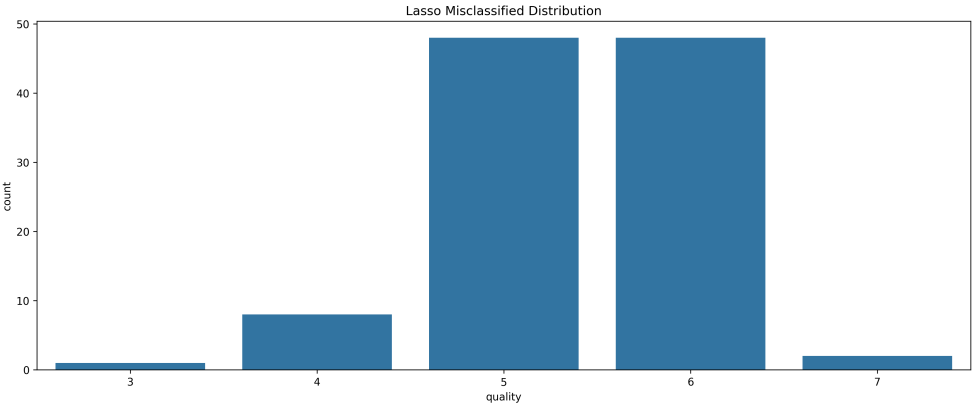
# Lasso Logistic Regression Results

The model accuracy is 0.732

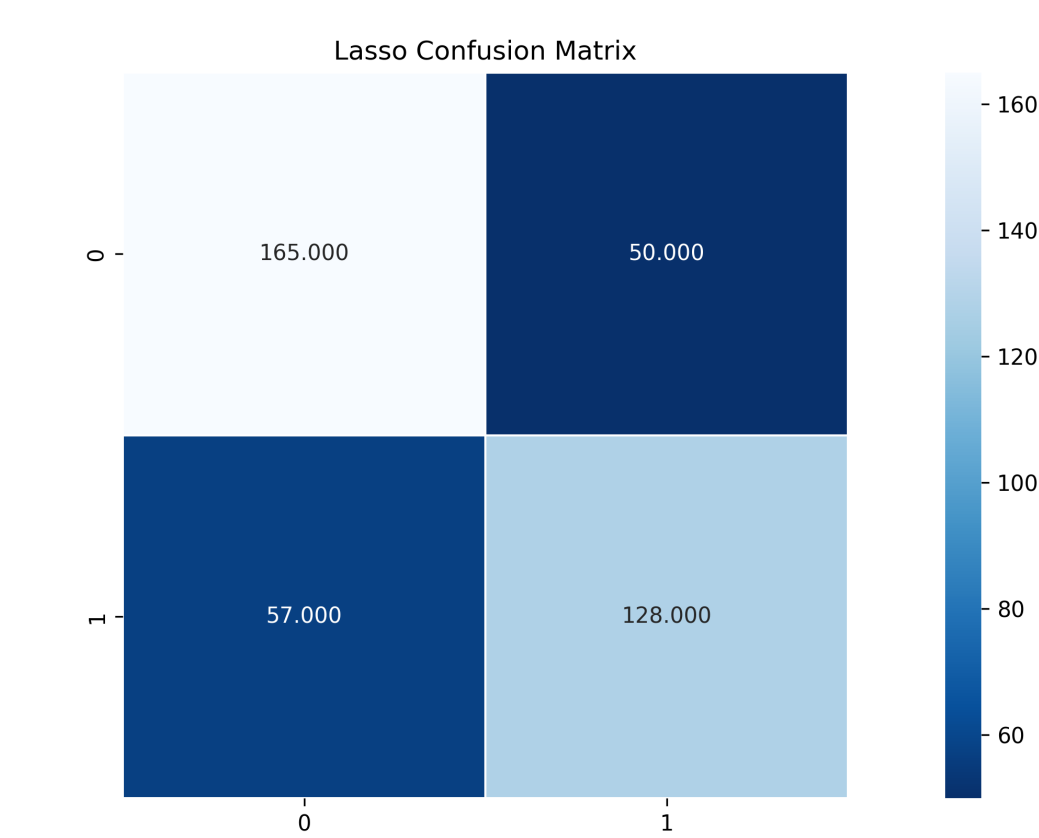
The model precision is 0.731

The model precision is 0.73

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 0.062

Feature volatile acidity has a coefficient 2.847

Feature citric acid has a coefficient 0.312

Feature residual sugar has a coefficient -0.01

Feature free sulfur dioxide has a coefficient -0.025

Feature total sulfur dioxide has a coefficient 0.019

Feature pH has a coefficient 2.005

Feature sulphates has a coefficient -0.996

Feature alcohol has a coefficient -0.842

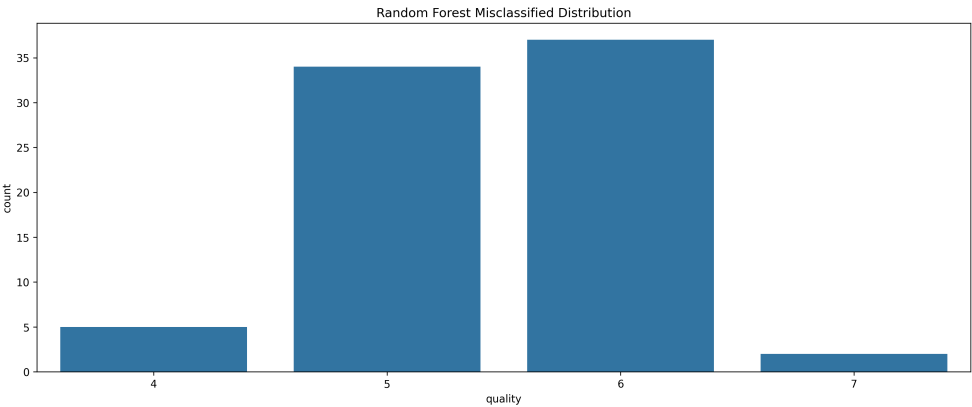
# Random Forest Results

The model accuracy is 0.805

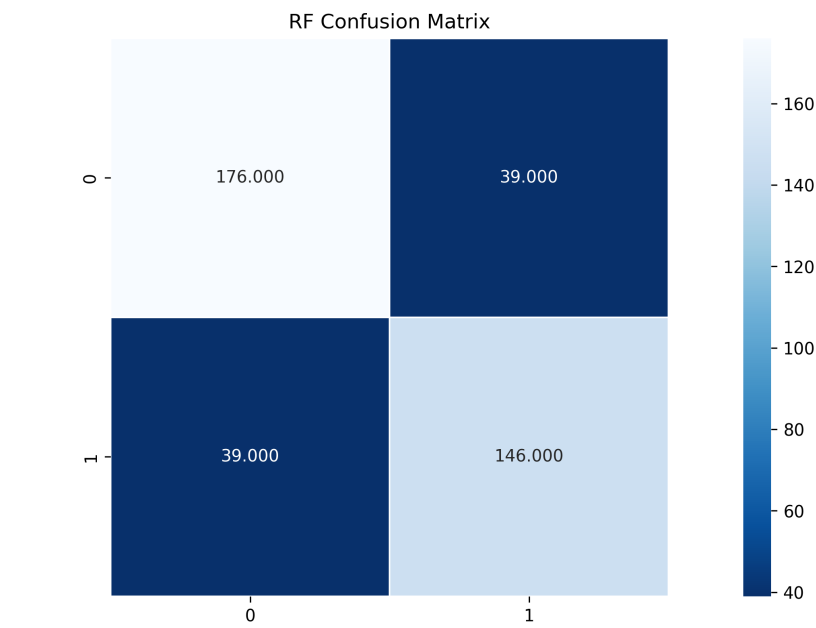
The model precision is 0.804

The model precision is 0.804

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



Below is a graph representing the confusion matrix



# Random Forest Drivers

