DATA 319

" Predictive Analysis of Wine Quality Through Physicochemical Properties "

Group 9 Project Members:

Halina Kuczynski (11786333)

Jenny Cheng (11678647)

Kyle Risso (11773294)

AGENDA

DATASET

https://archive.ics.uci.edu/dataset/186/wine+quality

Our group embarked on a comprehensive analysis of the "Wine Quality" dataset obtained from UC Irvine, with the primary goal of determining whether the physicochemical properties of wines could be used to predict and understand their quality.
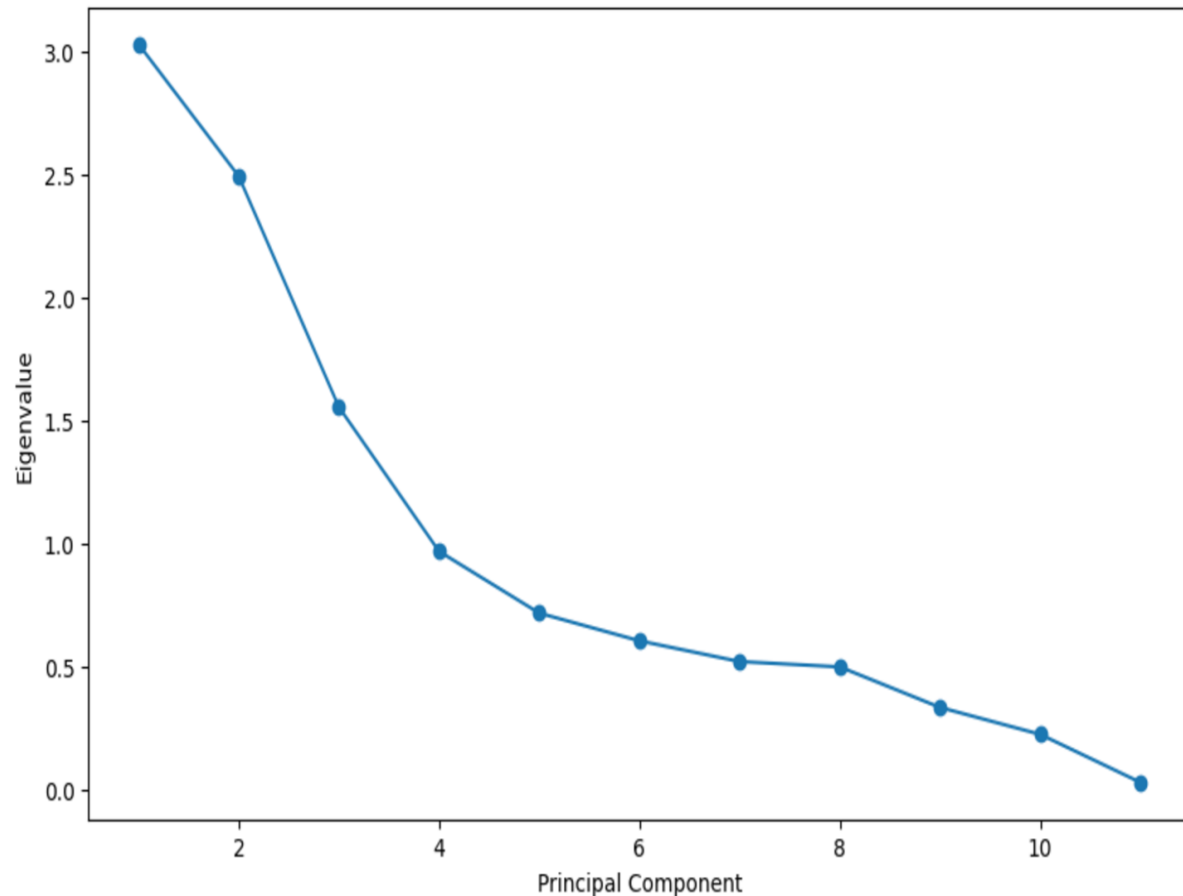
RESEARCH QUESTION

"Can we feasibly predict and understand the quality of wines based on their physicochemical properties?"
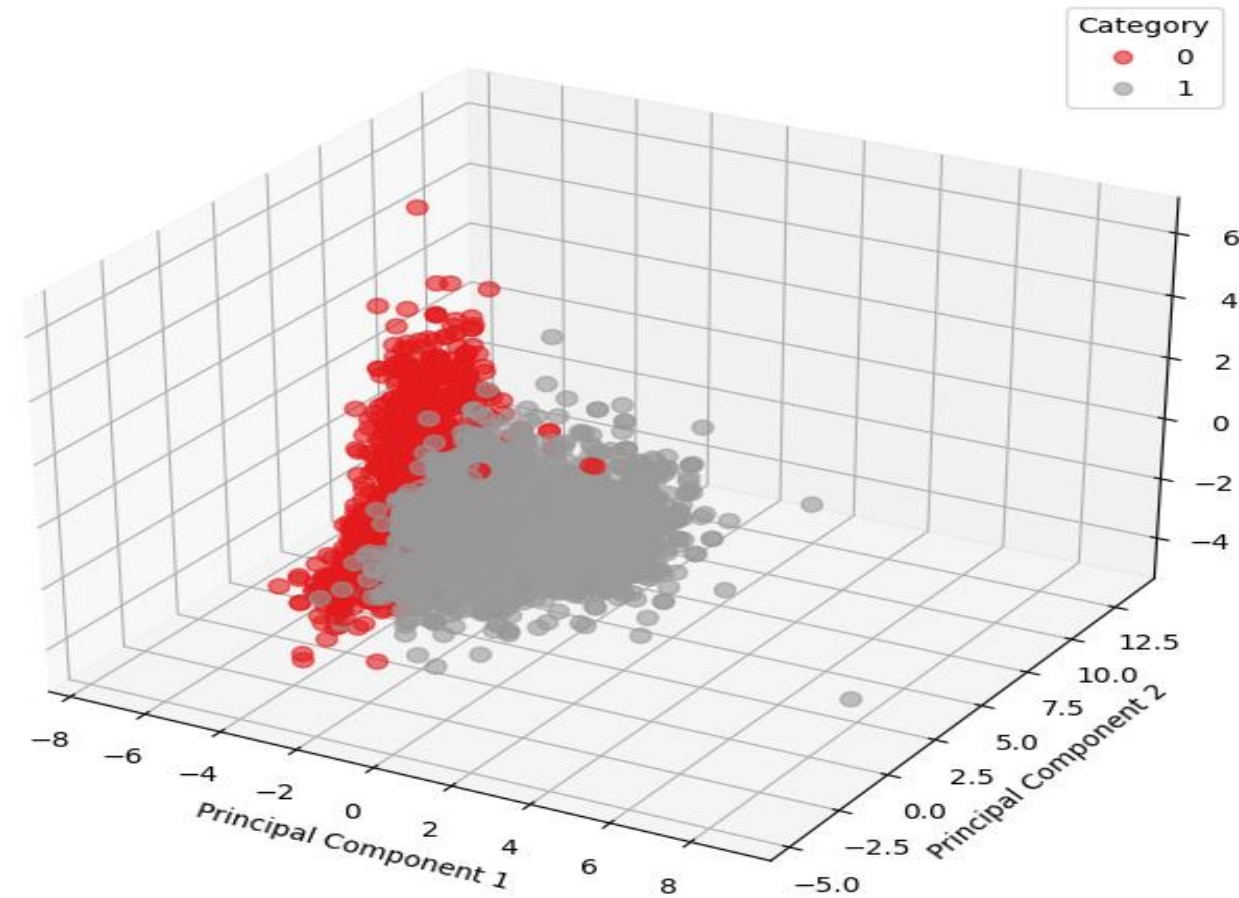
"Can multivariate analysis methods help identify patterns or clusters within the dataset that correlate with high or low-quality wines?"
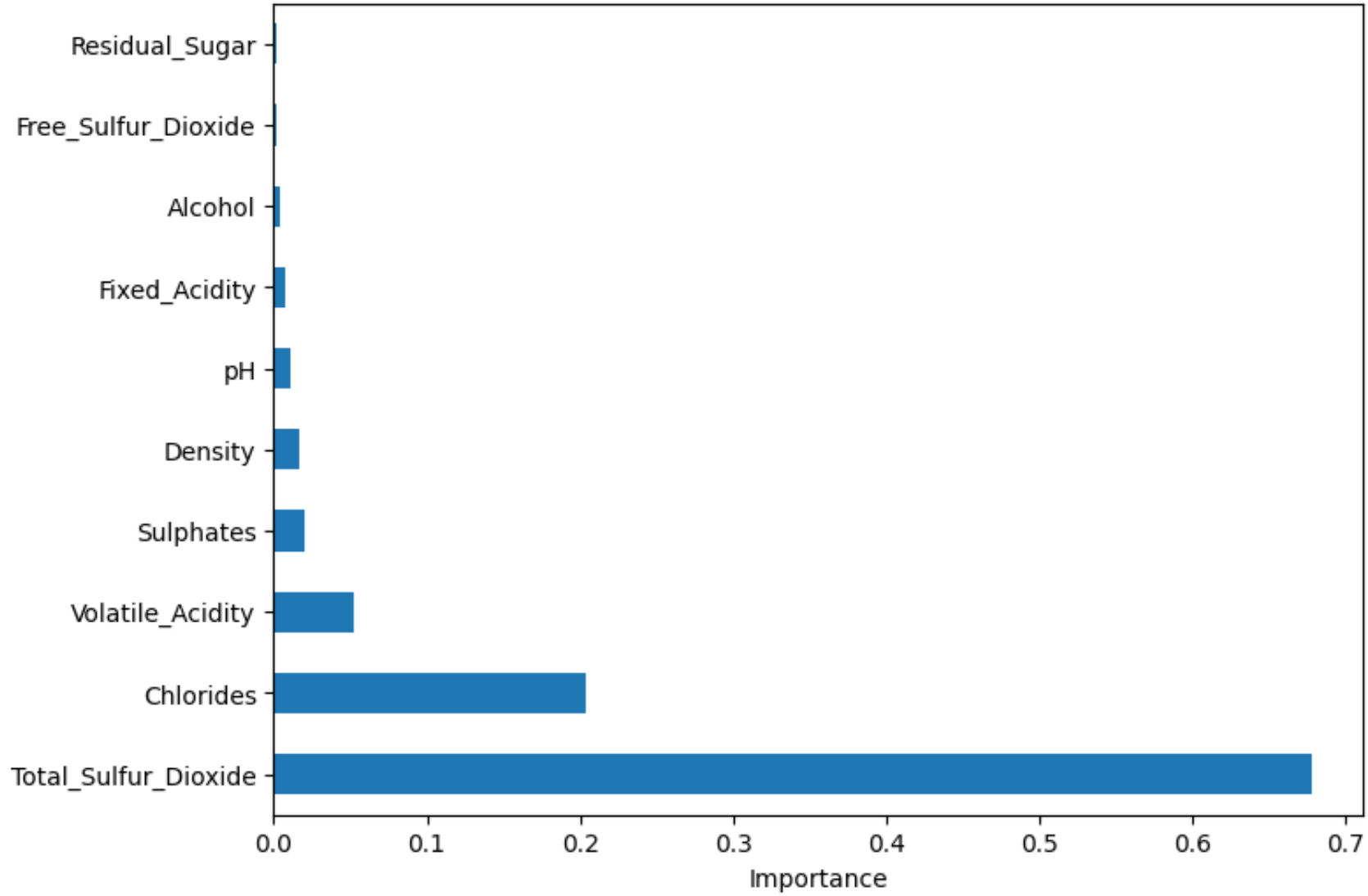
DIMENSION REDUCTION (PCA)

Scree Plot

PCA: 3D Scatter Plot of First Three Principal Components

Top 10 Feature Importances

CONFUSION HEATMAP

❖ 292 observations were correctly predicted as RED

❖ 7 observations were incorrectly predicted as RED

❖ 15 observations were incorrectly predicted as WHITE

❖ 986 observations were correctly predicted as WHITE

-> The model has a precision of 0.95 for RED, meaning that 95% of the RED predictions were correct. For WHITE, it's even higher at 0.99.
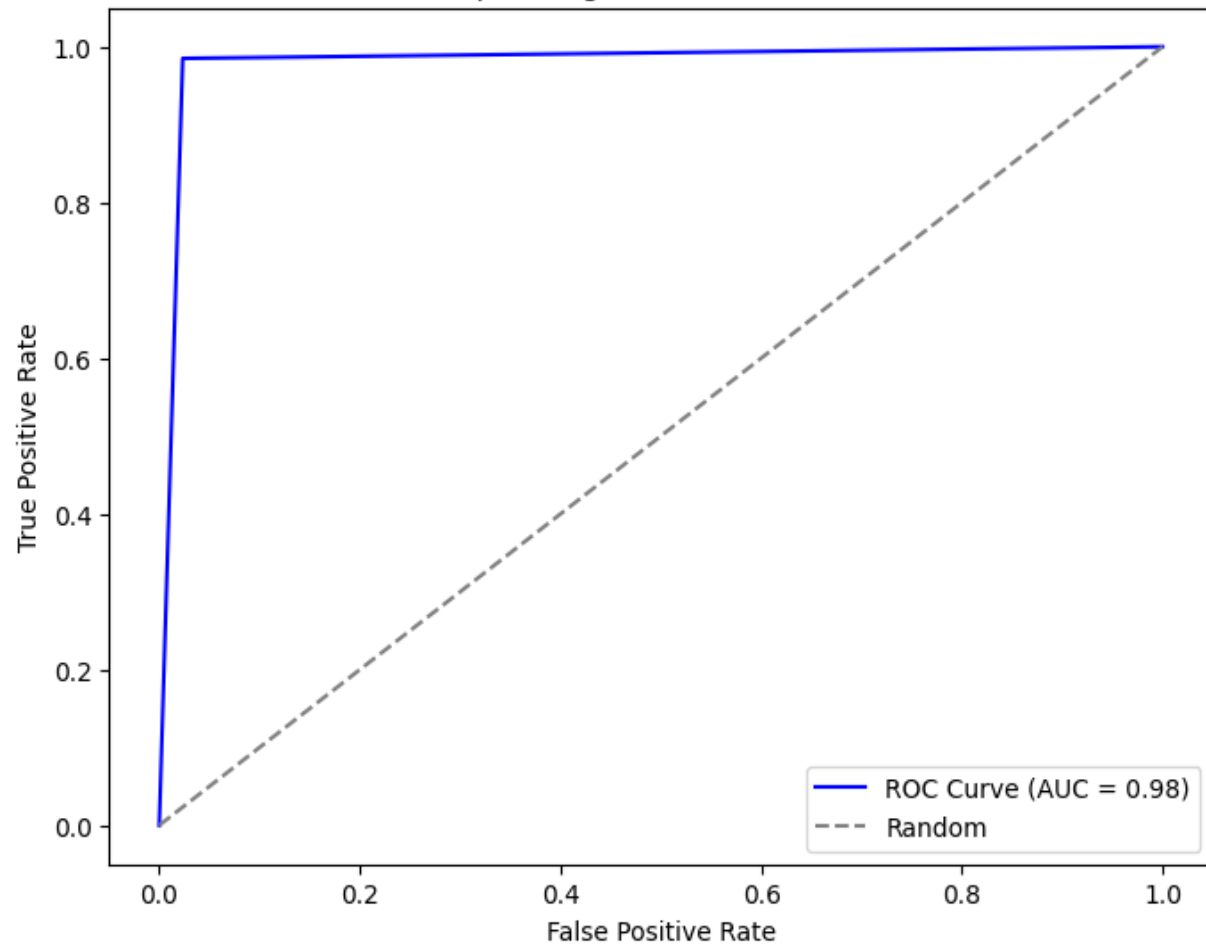
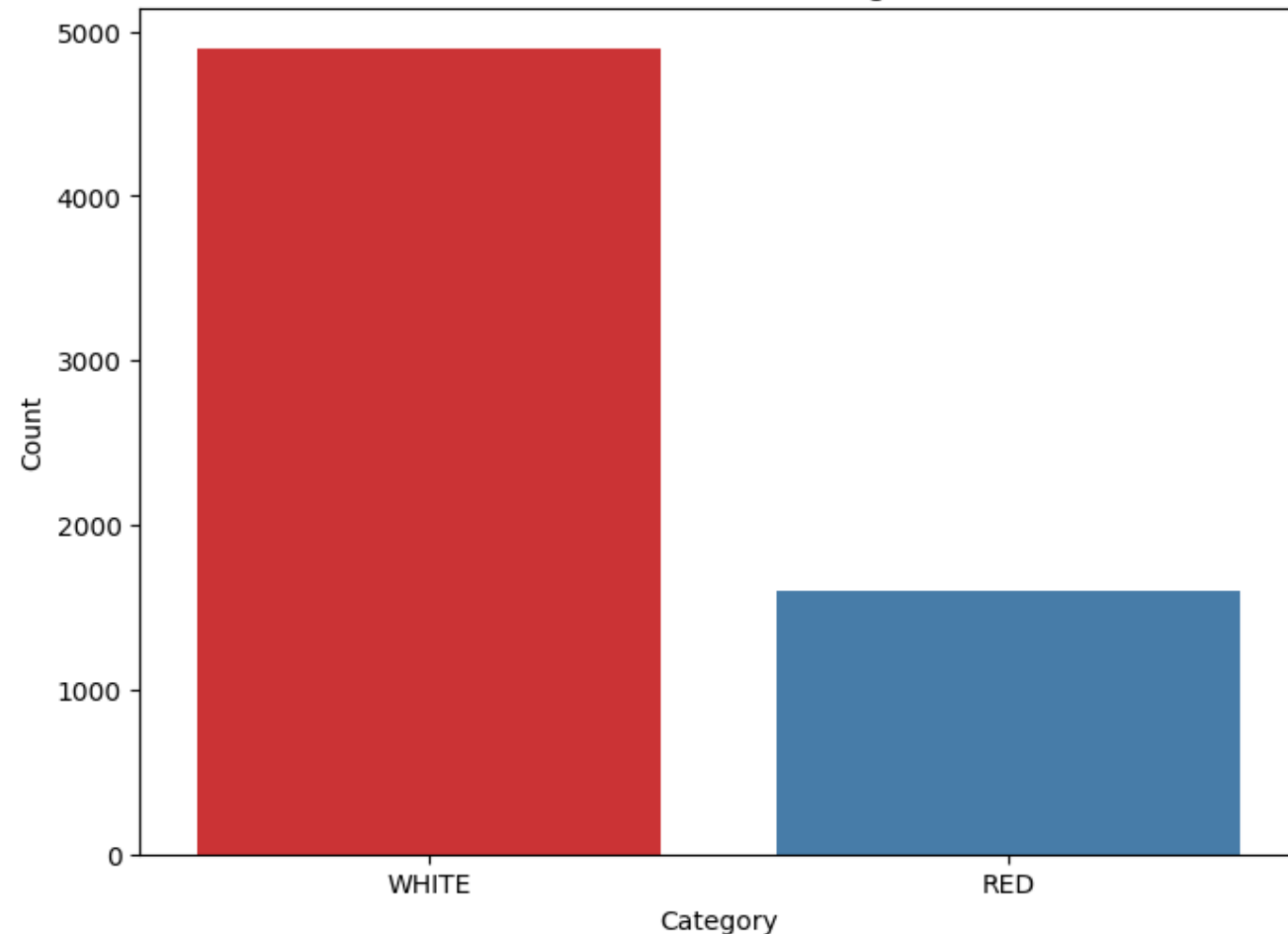-> The F-1 is 0.96 for RED and 0.99 for WHITE, which are both high, indicating good performance.

Confusion Matrix
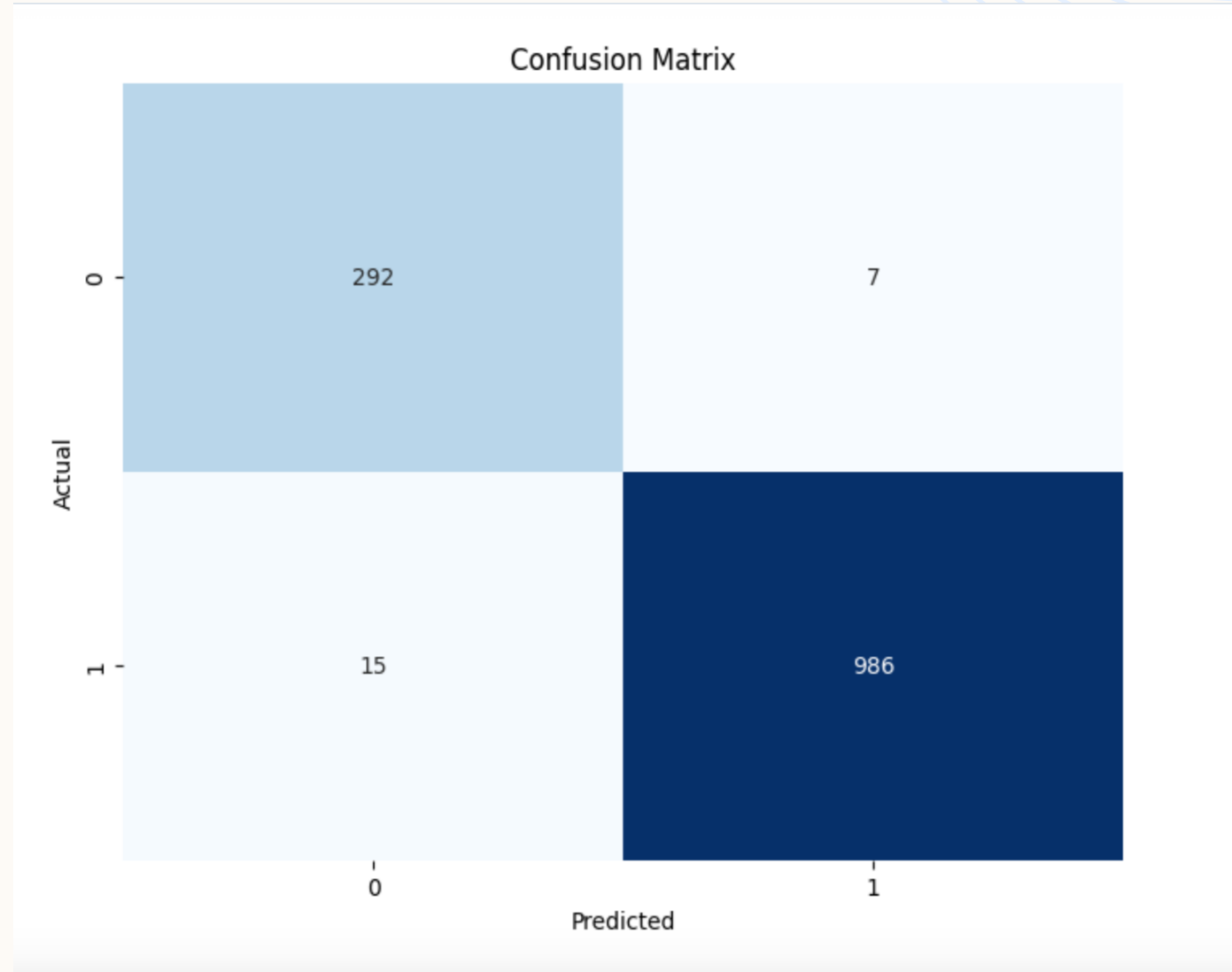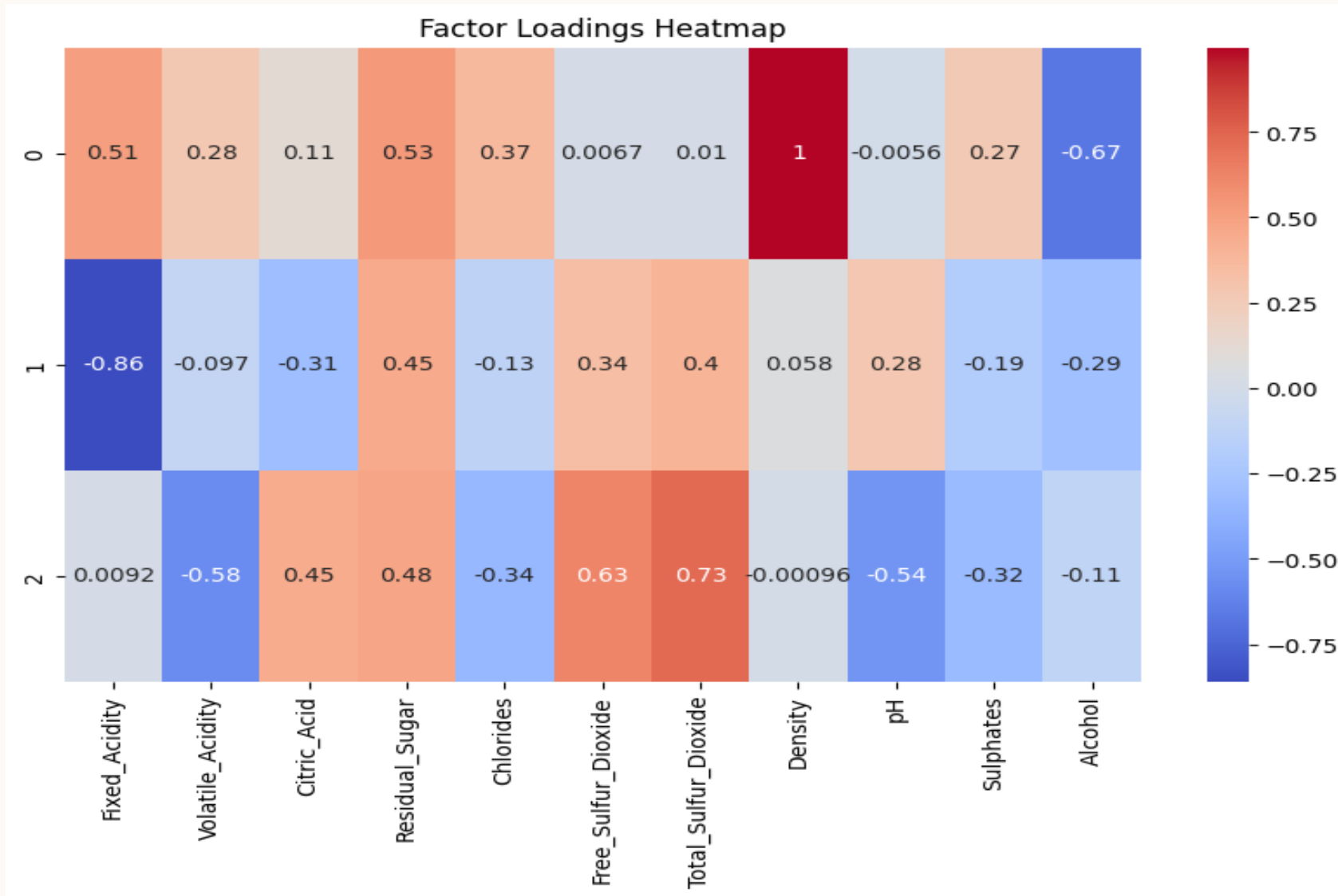
ROC CURVE

Receiver Operating Characteristic (ROC) Curve

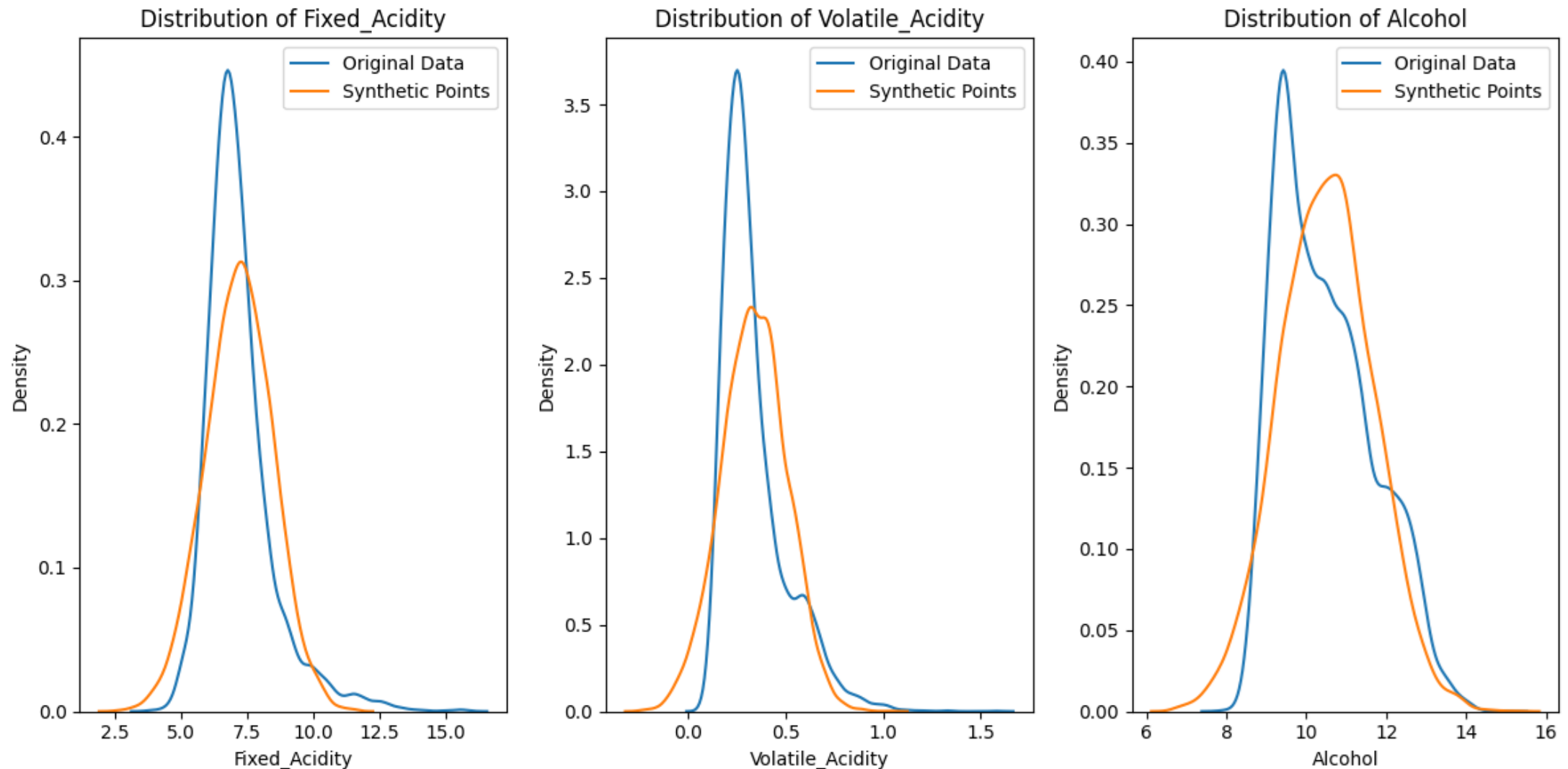Distribution of Wine Categories

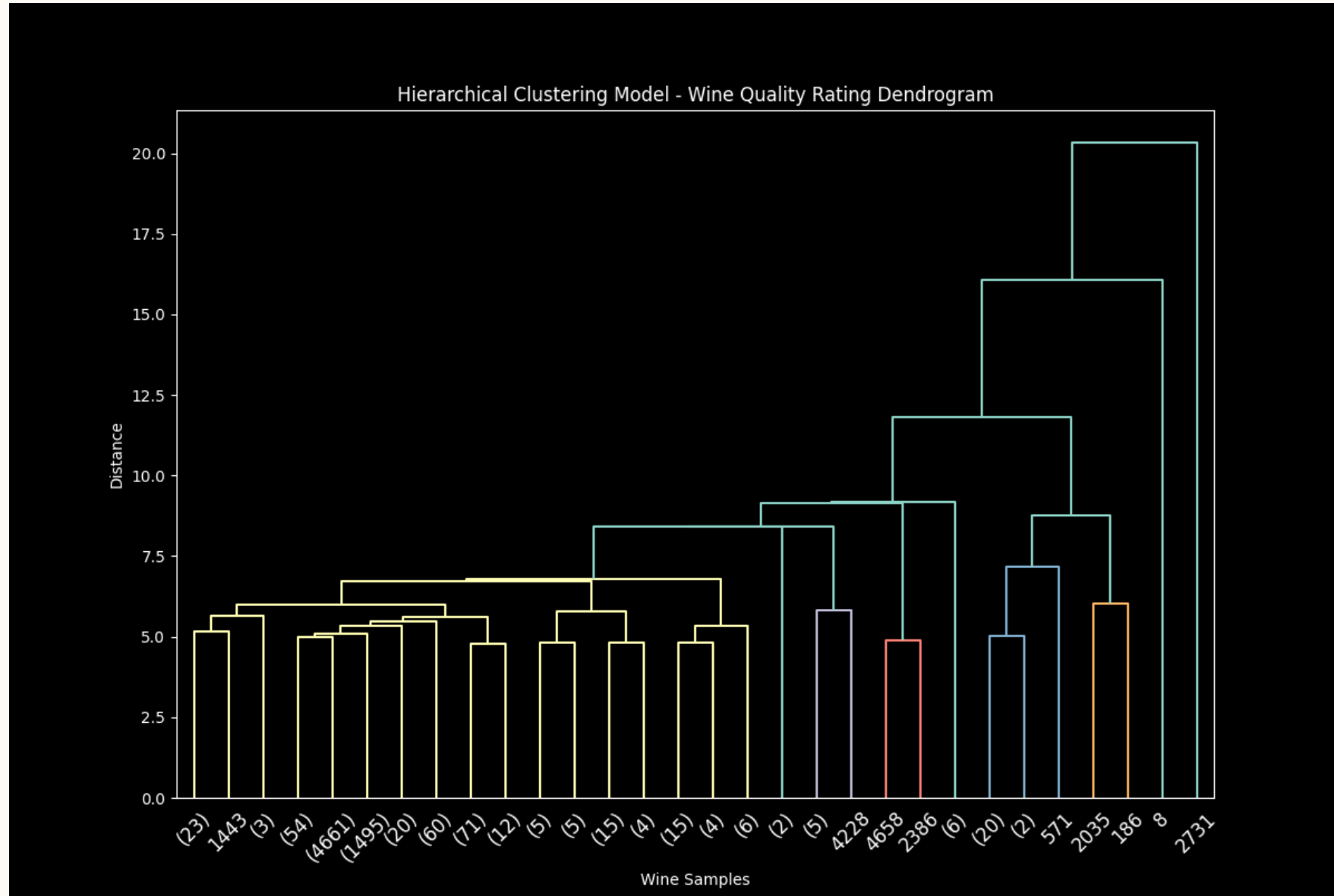# FACTOR ANALYSIS

Factor Loadings Heatmap

MULTIVARIATE NORMAL MODELING

HIERARCHICAL CLUSTERING

CONCLUSION & OBSERVATIONS

Ans 1:

Yes, the quality of wines can often be predicted and understood based on their physicochemical properties, which is a common practice in the field of chemometrics.

Ans 2:

Absolutely, multivariate analysis methods are well-suited for identifying patterns or clusters within a dataset that correlate with high or low-quality wines. These methods allow for the simultaneous analysis of multiple variables, providing a more comprehensive understanding of the relationships and structures within the data.

-> Two key multivariate analysis techniques that can be particularly useful for identifying patterns or clusters are Principal Component Analysis (PCA) and clustering algorithms.

THANK YOU!