CLUSTERING PROJECT:

Predicting the Quality of Wine

Annie Carter
July 23, 2023

data.world

Agenda

WHAT MAKES A QUALITY WINE?

EXECUTIVE SUMMARY

FINDINGS

CONCLUSION

Executive Summary

GOALS

Explore wine quality predictors, analyze chemical property clusters using KMeans, assess the predictive performance of regression models, and employ the selected model for future quality predictions.

BIG IDEA

. The key predictors associated with decreased quality include chlorides, density, and volatile acidity. Improving these predictors leads to an enhancement in wine quality.

FINDINGS

- White wine rated higher than red
- Notable predictors of quality: chlorides, density, and volatile acidity
- Cluster analysis insufficient
- Regression Lasso+Lars model beat baseline by 17%

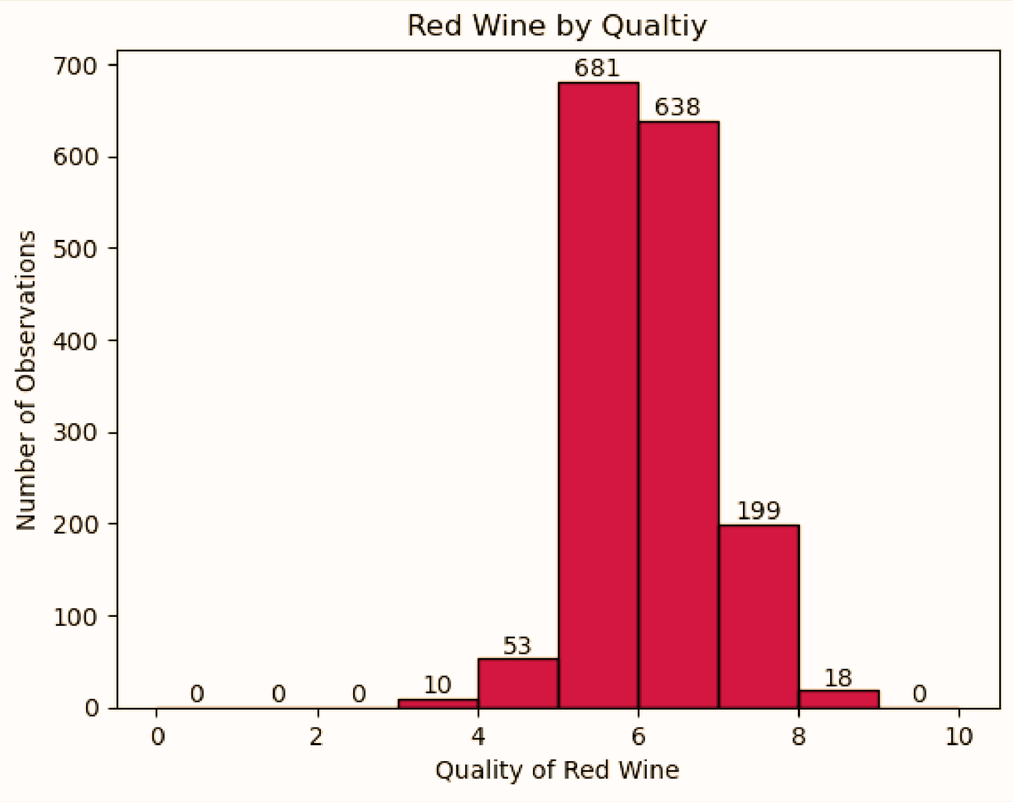
RECOMMENDATIONS

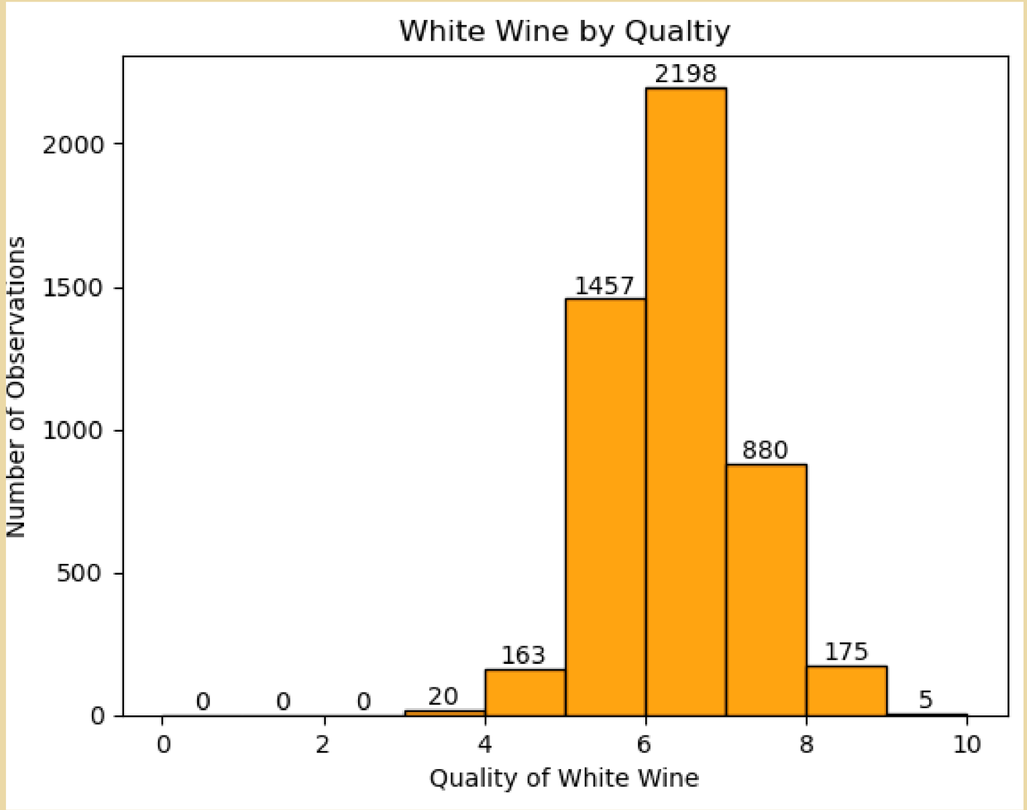
- Conduct separate evaluations for white and red wines
- Involve more wine experts for more data
- Use DBSCAN cluster analysis
- Collect Data on additional features (e.g., citric acid, residual sugar)

Wine Quality Findings

DOES THE TYPE OF WINE "RED OR WHITE" HAVE A RELATIONSHIP TO WINE QUALITY?

- White wine is statistically rated higher in quality than red wine, with a difference of approximately 0.3 points. White wines are lighter and fruitier making them ideal for everyday refreshment
- Compared to red wines, white wines have lower tannin levels than red wines. Tannins are compounds that can give red wine a dry, tangy taste. Which some may not prefer.





Wine Quality Findings

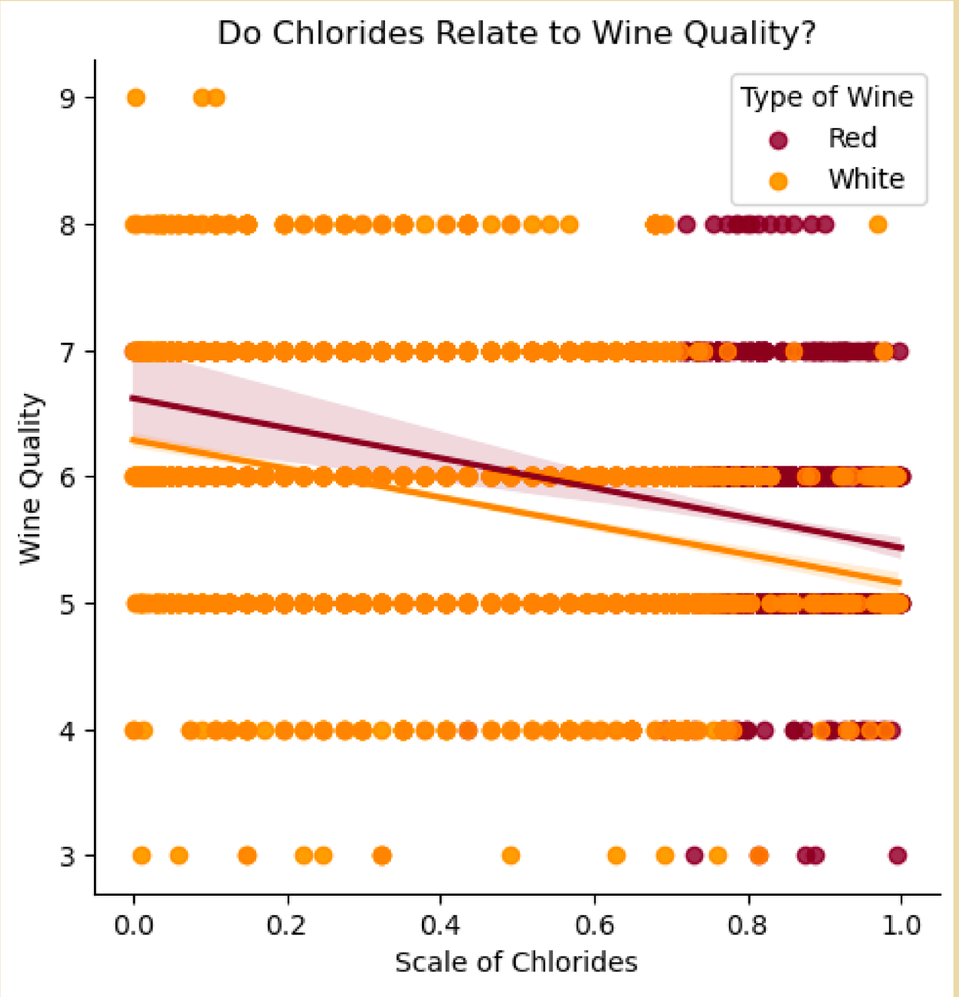
Notable Features that Predict Wine

DOES CHLORIDE HAVE A RELATIONSHIP TO WINE QUALITY ?

Data visualization and statistic testing reveals a relationship between Chlorides and wine quality. Chloride content in wine is influenced by terroir (environment) and grape variety(aroma, acidity, tannin). impacting flavor. High concentrations may result in an undesirable salty taste, affecting market appeal (Coli et al., 2015).

Volatile Acidity and Density had similar relationships where increase in feature led to decrease in quality

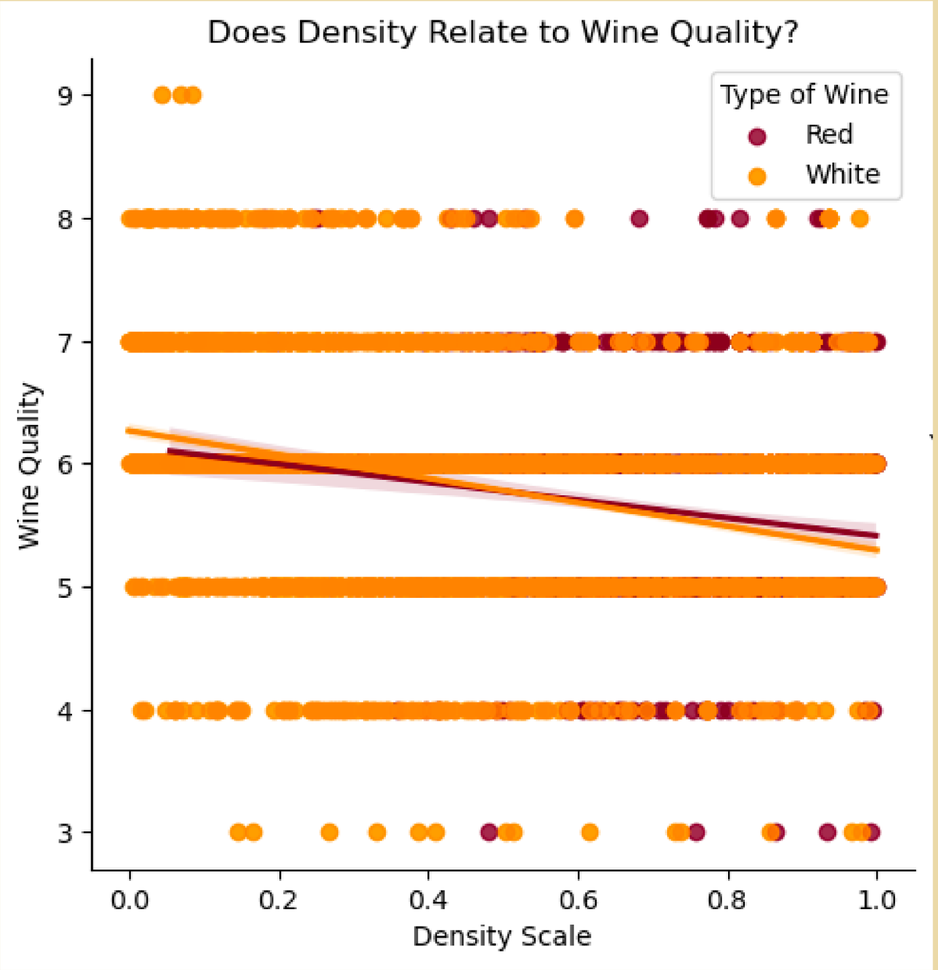
Coli, M. S., Rangel, A. G. P., Souza, E. S., Oliveira, M. F., & Chiaradia, A. C. N. (2015). Chloride concentration in red wines: influence of terroir and grape type. Food Science and Technology, 35, 95-99.

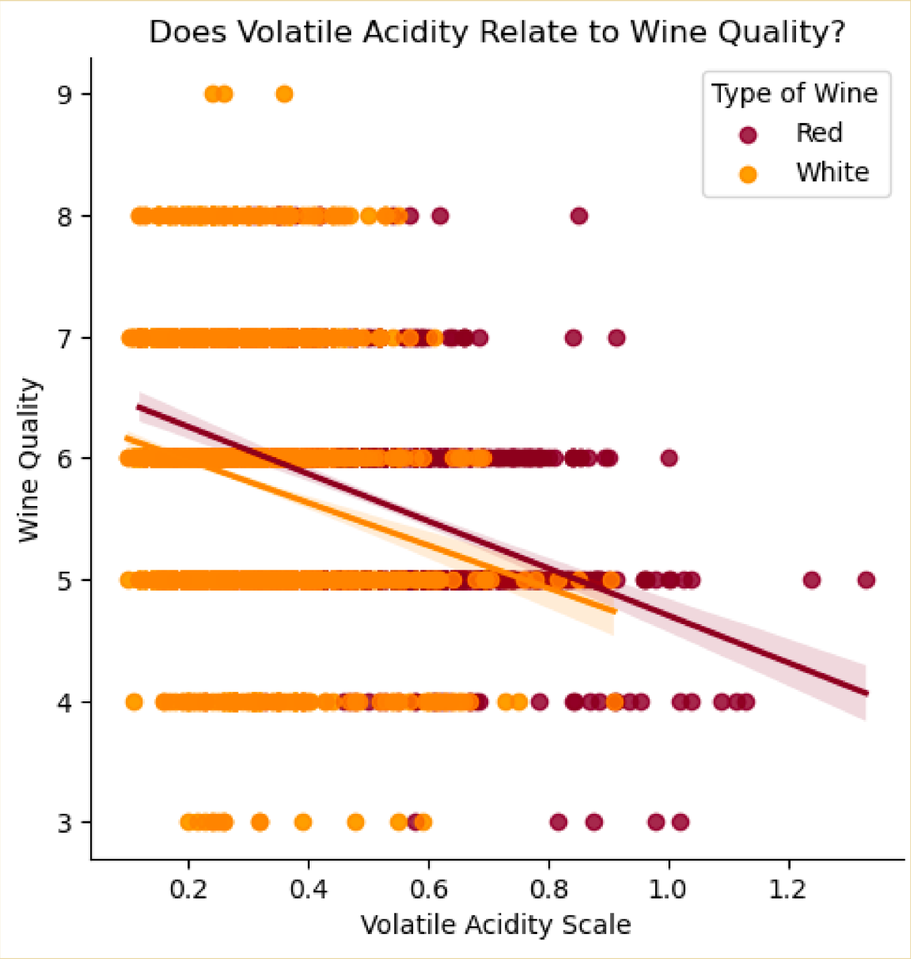


Wine Quality Findings

Notable Features that Predict Wine

DOES DENSITY OR VOLATILE ACIDITY HAVE A RELATIONSHIP TO WINE QUALITY ?





Wine Quality Findings

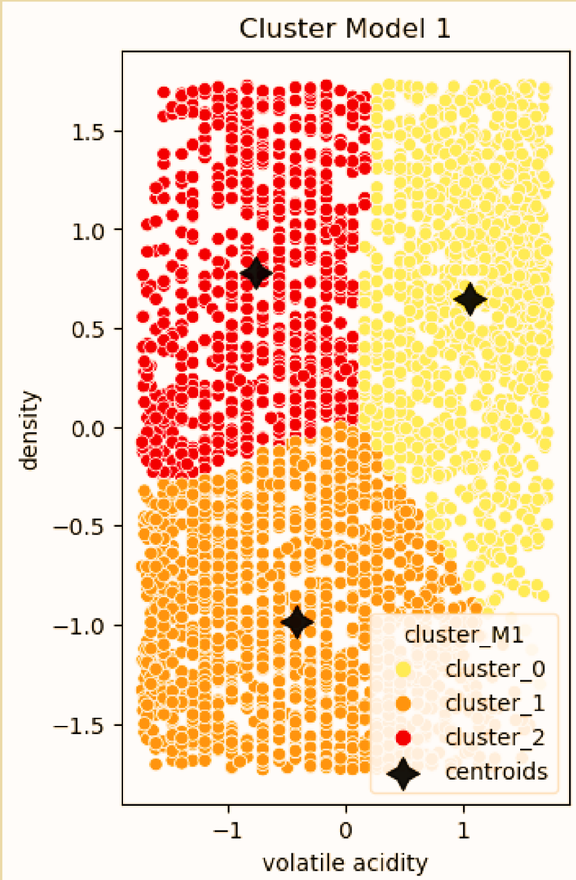
Cluster Analysis and Regression Modeling

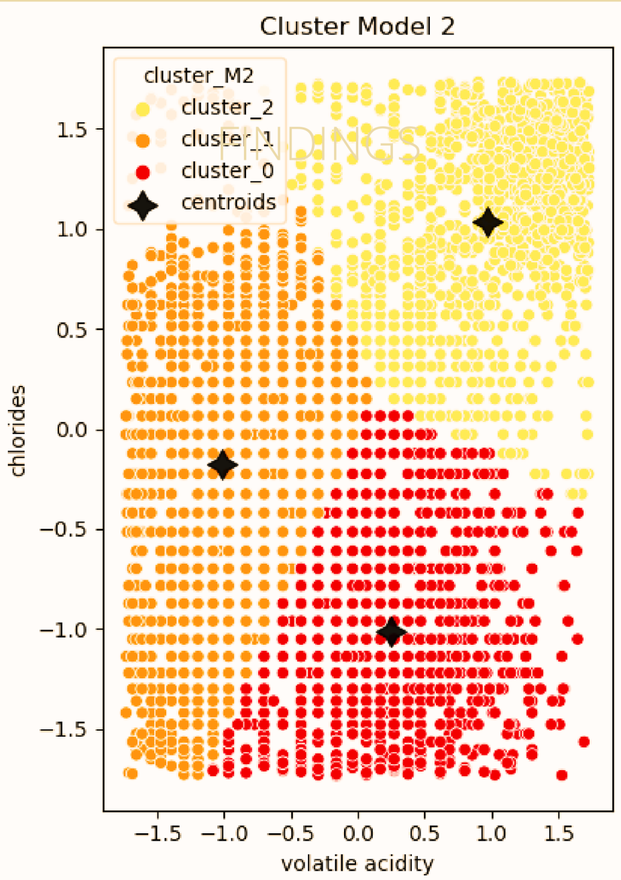
CLUSTER ANALYSIS

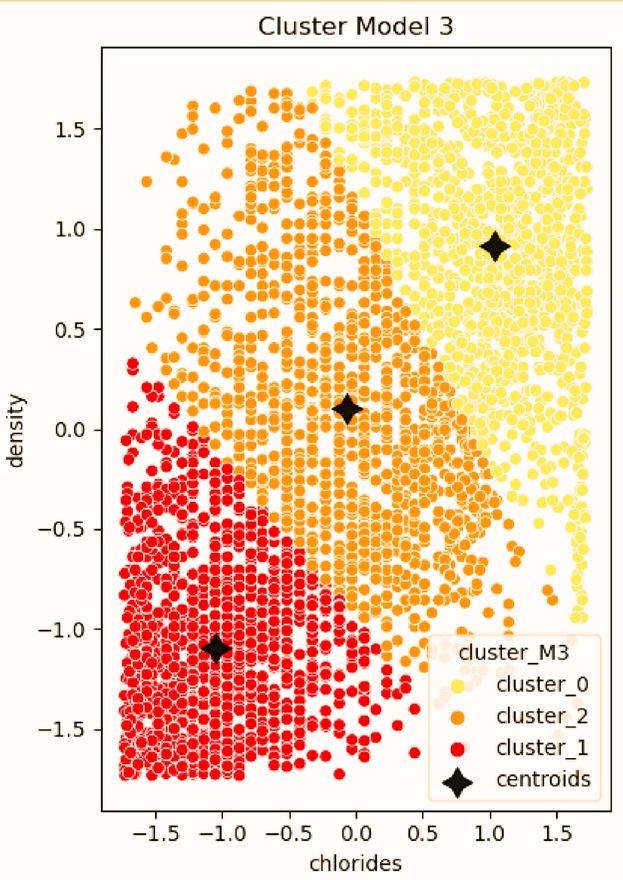
Cluster Analysis Results: Indistinct clusters made it unsuitable for regression modeling.

Regression Models Implemented: Four

Top Performer: Lasso+Lars consistently outperformed the baseline by 17%.







Conclusion

WHAT MAKES A QUALITY WINE?

This project used regression models and KMeans cluster analysis to identify value clusters. However, cluster analysis did not predict future wine quality effectively. White wine rated significantly higher than red wine. Key predictors for wine quality were volatile acidity, chlorides, and density, with an increase in feature having a negative impact on quality

Top Model: Among the regression models, LASSO + LARS stood out as the top performer, outperforming the baseline by 17%.

NEXT STEPS

- Consider conducting DBSCAN cluster analysis to eliminate outliers and potentially define unique clusters more effectively.

- Collect data on variables (e.g., temperature, duration of fermentation to see if they may improve data for future clustering

RECOMMENDATIONS

- To gain more nuanced insights, conducting separate evaluations for white and red wines is advised, with emphasis on minimizing chlorides (associated with saltiness) and volatile acidity (linked to vinegar flavor) for improved wine quality.

- Enhance data collection by including more wine experts (e.g., 4-6) to gather additional observations for analysis.

FOR MORE INFORMATION CONTACT

• ANNIE CARTER

•HTTPS://GITHUB.COM/ANNIE-CARTER/CLUSTERING_PROJECT


Thank You