

SOMMELIER SCIENCERS

May 30, 2023

PRESENTERS

Alexia Lewis

Exploration/Clustering

Oliver Taylor

Classification/Regression

AGENDA

Executive Summary

Findings

Recommendations

Conclusion

EXECUTIVE SUMMARY

Big Idea

Using ML techniques we can take the chemicals of wine and predict the finished wine's quality similar to a sommelier

Top Findings

Acidity, sodium, and alcohol levels affect quality

Quality Rating is 3-9

Goal

Predict the Quality of Vinho Verde Wine

Recommendation

Future Researchers: Gather data on more grape types.

Data Engineers: Either split the dataset into white and red and create separate models or gather more data on red types of vinho verde to balance the dataframe and provide additional insights.

Data Scientists: Remove outliers and engineer "acid" and "feel" features using appropriate columns.

Business: We do not recommend putting the regression models into production.

Moving into the Executive Summary

EXECUTIVE SUMMARY

Big Idea

Using ML techniques we can take the chemicals of wine and predict the finished wine's quality similar to a sommelier

Top Findings

Acidity, sodium, and alcohol levels affect quality

Quality Rating is 3-9

Goal

Predict the Quality of Vinho Verde Wine

Recommendation

Future Researchers: Gather data on more grape types.

Data Engineers: Either split the dataset into white and red and create separate models or gather more data on red types of vinho verde to balance the dataframe and provide additional insights.

Data Scientists: Remove outliers and engineer "acid" and "feel" features using appropriate columns.

Business: We do not recommend putting the regression models into production.

We have the big idea which states that

By using both clustering and classification modeling we can take the raw chemical makeup of wine and generate wine quality predictions similar to that of a trained sommelier.

EXECUTIVE SUMMARY

Big Idea

Using ML techniques we can take the chemicals of wine and predict the finished wine's quality similar to a sommelier

Top Findings

Acidity, sodium, and alcohol levels affect quality

Quality Rating is 3-9

Goal

Predict the Quality of Vinho Verde Wine

Recommendation

Future Researchers: Gather data on more grape types.

Data Engineers: Either split the dataset into white and red and create separate models or gather more data on red types of vinho verde to balance the dataframe and provide additional insights.

Data Scientists: Remove outliers and engineer "acid" and "feel" features using appropriate columns.

Business: We do not recommend putting the regression models into production.

Our top findings were that acidity, sodium and alcohol levels affect the quality of wine.

We also found that the rating scale is actually 3 through 9, not 0 through 10.

EXECUTIVE SUMMARY

Big Idea

Using ML techniques we can take the chemicals of wine and predict the finished wine's quality similar to a sommelier

Top Findings

Acidity, sodium, and alcohol levels affect quality

Quality Rating is 3-9

Goal

Predict the Quality of Vinho Verde Wine

Recommendation

Future Researchers: Gather data on more grape types.

Data Engineers: Either split the dataset into white and red and create separate models or gather more data on red types of vinho verde to balance the dataframe and provide additional insights.

Data Scientists: Remove outliers and engineer "acid" and "feel" features using appropriate columns.

Business: We do not recommend putting the regression models into production.

The goal of this project is to predict the quality of wine.

EXECUTIVE SUMMARY

Big Idea

Using ML techniques we can take the chemicals of wine and predict the finished wine's quality similar to a sommelier

Top Findings

Acidity, sodium, and alcohol levels affect quality

Quality Rating is 3-9

Goal

Predict the Quality of Vinho Verde Wine

Recommendation

Future Researchers: Gather data on more grape types.

Data Engineers: Either split the dataset into white and red and create separate models or gather more data on red types of vinho verde to balance the dataframe and provide additional insights.

Data Scientists: Remove outliers and engineer "acid" and "feel" features using appropriate columns.

Business: We do not recommend putting the regression models into production.

Recommendations we can dive deeper into later but for now, we need more data and we don't recommend business action at this time.

EXPLORATION FINDINGS

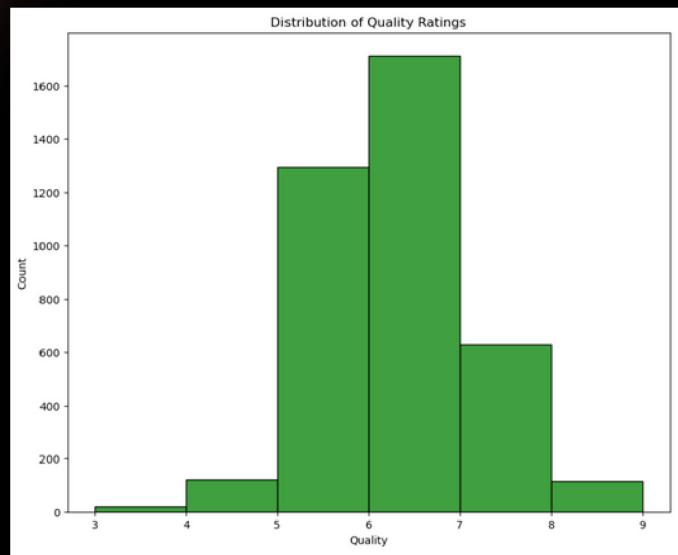
- Acidity, sodium, and alcohol affect quality
- White wine results in a much wider range of residual sugars
- Type has a 75/25 percent proportion
 - White = 75%
 - Red = 25%
- pH typical range 0-14
 - wine range 2.7 - 4
- Quality Rating is 3-9

During our exploration phase we uncovered that acidity, alcohol, and sodium levels affect the quality rating.

75% of the data is for white wine and 25% is for red wine and the pH lies between 2.7 and 4.

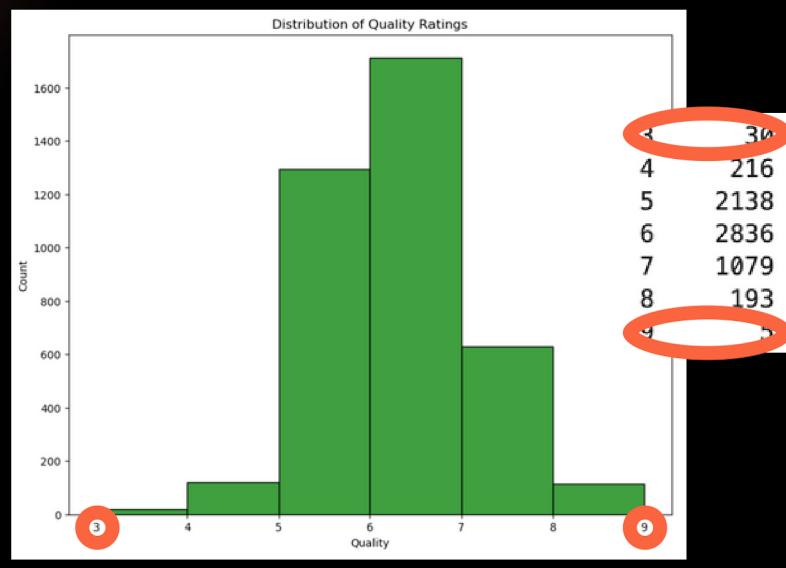
For context, battery acid is a 1, vinegar is a 3, and tomato juice is a 4.

EXPLORATION FINDINGS



We also discovered that the quality rating does not go from 0 to 10 like originally thought.

EXPLORATION FINDINGS



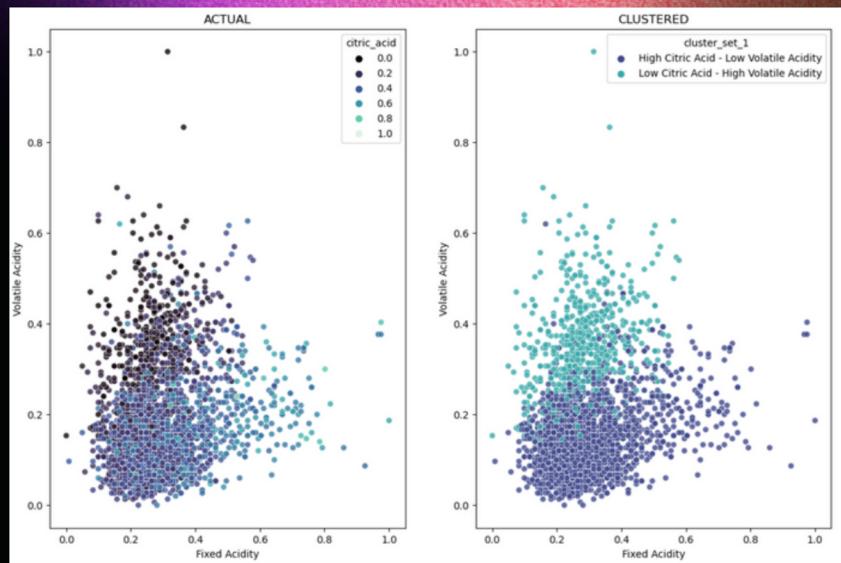
The quality range is actually 3-9.

Wines with a quality of 3 have a count of 30 wines.

Those with a quality of 9 have a count of 5 wines.

The mean is 6

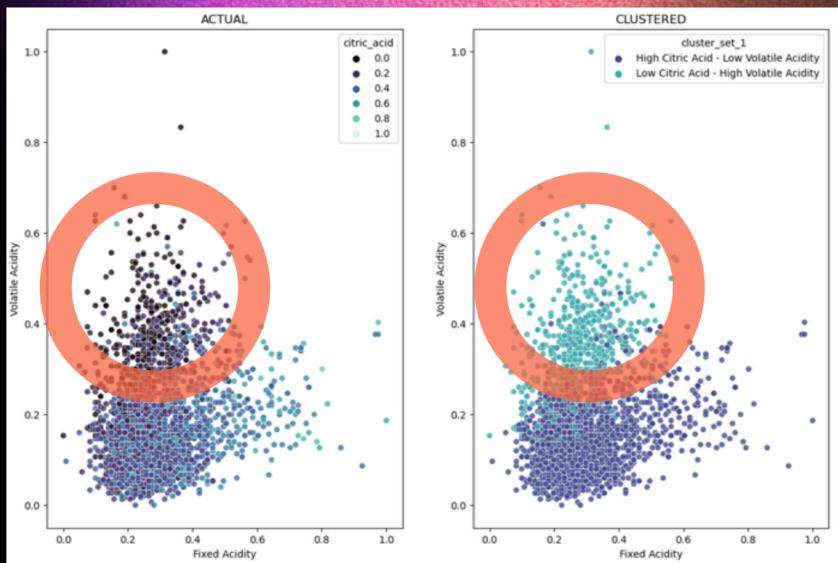
CLUSTERING FINDINGS



We learned that volatile acidity results from the degradation of citric acid. This type of acidity can give wine a more vinegar-like flavor. Since these columns are related, we engineered a feature using clustering called "acid."

The graph on the left is the actual data and the graph on the right shows the clustered data. While there aren't any clear clusters, it did help identify the relationship between acid-related data.

CLUSTERING FINDINGS



- Volatile acidity results from the degradation of citric acid
- The actual relationship and the clustered relationship created the feature "Acid"

The darker dots on the left indicate lower citric acid and higher volatile acidity, which checks against our domain knowledge that states volatile acid results from the degradation of citric acid.

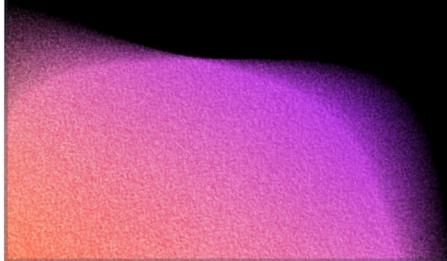
These darker dots also correspond with the lighter dots on the right graph displaying the clustering results, indicating a relationship.

We used this engineered feature in our model.

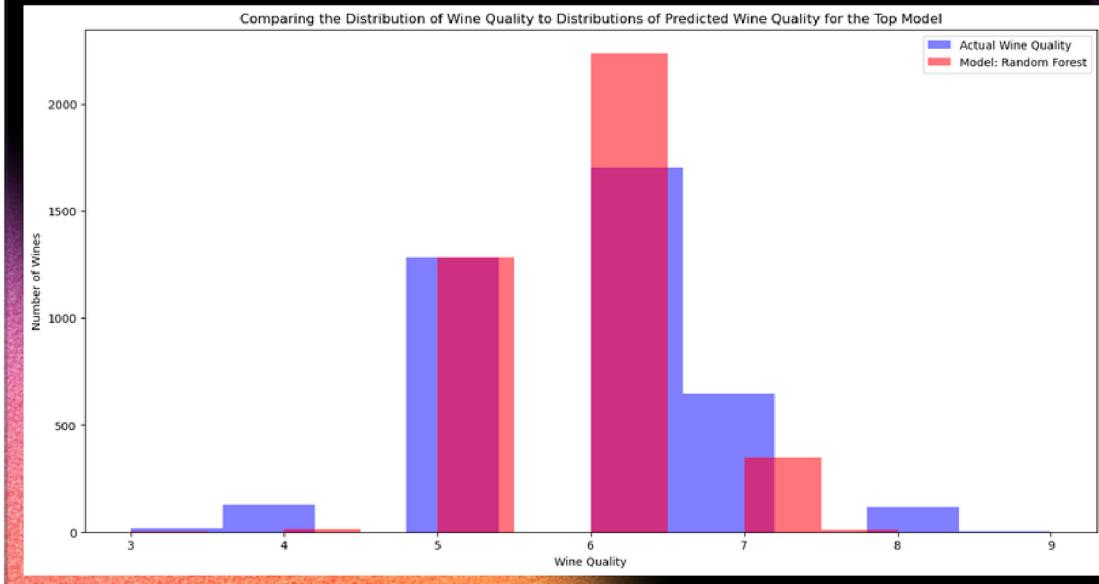
MODELING KEY FINDINGS

Several classification and regression models were validated and the random forest classification model at a depth of 8 was the best performing model

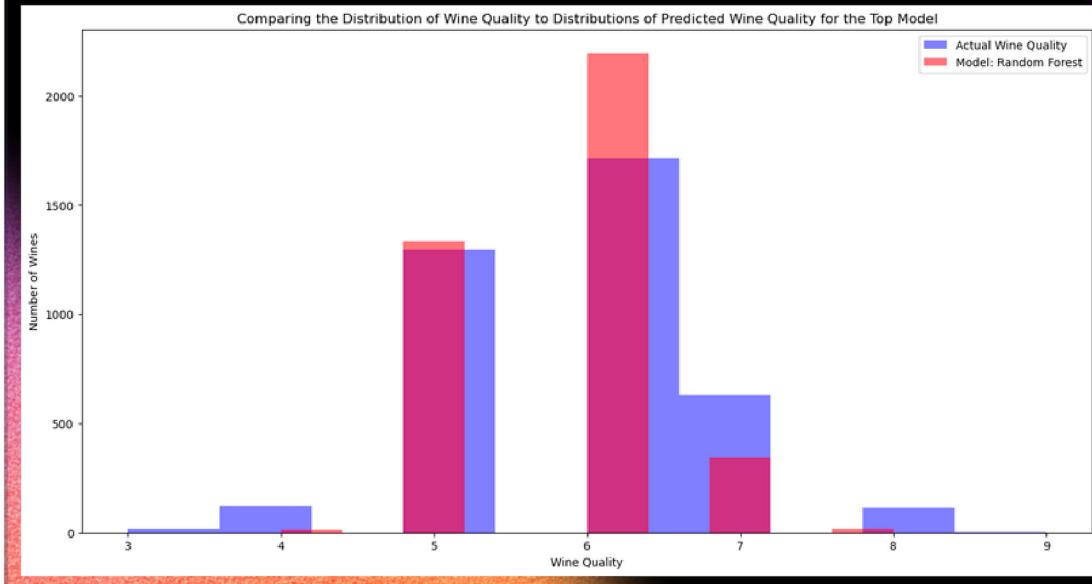
The model beat the baseline of 44% with a accuracy score of 59%



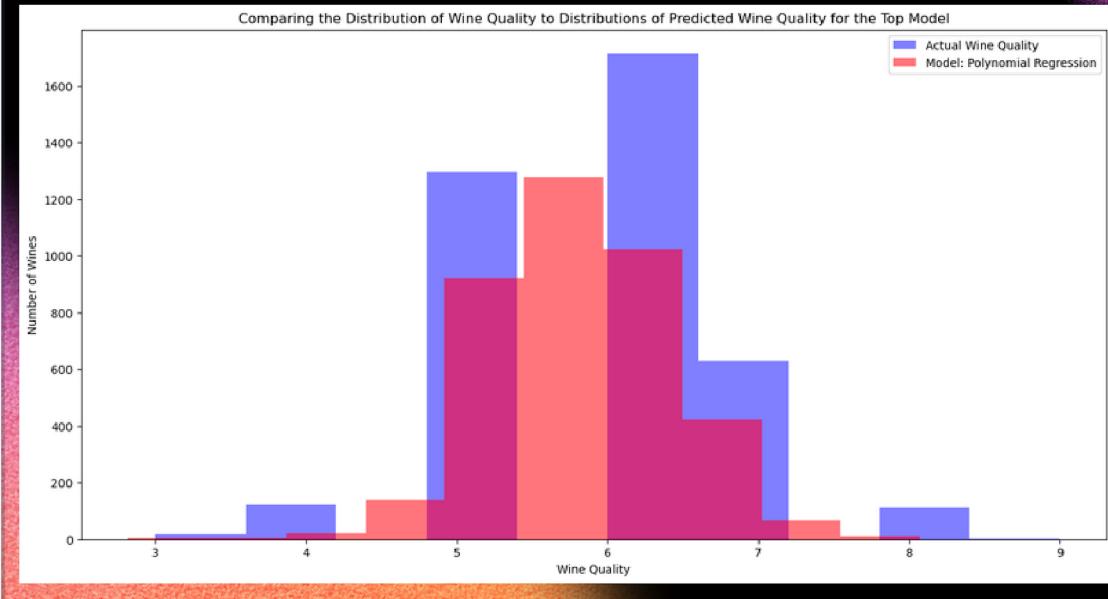
MODELING FINDINGS



MODELING FINDINGS

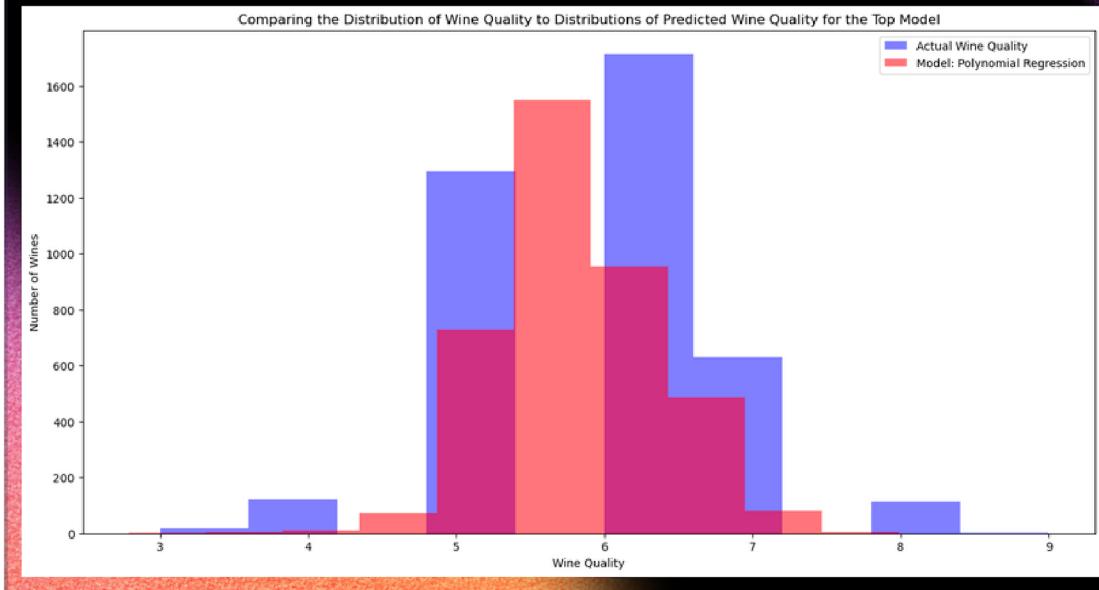


MODELING FINDINGS



Polynomial Model
without clusters
Baseline RMSE: .87
Test RMSE: .72
Test: r^2 : .31

MODELING FINDINGS



Polynomial Model
without clusters
Baseline RMSE: .87
Test RMSE: .72
Test r²: .31

CONCLUSION

- 3 key drivers are acidity, sodium, and alcohol content
- Clustering identified relationships which we feature-engineered and modeled.
The feature from clustering ("acid") did not significantly affect the predictive modeling outcomes.
- The classification model, random forest, produced the best results with decent precision numbers and an accuracy score of 59% beating the baseline of 44%.
- Out of the regression models the polynomial model performed at an RMSE of .72.

KEY CONCLUSION TAKEAWAY

We **CAN** predict wine quality using cluster and predictive modeling and get outcomes similar to those of a sommelier.

RECOMMENDATIONS AND NEXT STEPS

- **Future Researchers:** Gather data on more grape types.
- **Data Engineers:** Either split the dataset into white and red and create separate models or gather more data on red types of vinho verde to balance the dataframe and provide additional insights.
- **Data Scientists:** Remove outliers and engineer "acid" and "feel" features using appropriate columns.
- **Business:** The models succeed as a proof of concept but we do not recommend putting these models into production at this time. With further research more accurate models could be made.

THE END

FOR SOMMELIERS

The age of the robot sommelier is upon us.