••••

By: Alicia Gong, Ashley Chen, Abdel Shehata, Claire Tan

Introduction To Topic

Facts of the matter

About 234 million hectoliters of wine were consumed in 2020.

Vinho Verde region is one of the largest and oldest wine regions in the world.

There are different chemical factors that affect the quality of Vinho Verde wine.

Possible Reasons

➔ Variation in Acidity
➔ Variation in Sugar content
➔ Variation in PH
➔ Variation in Alcohol content

Research Question

What chemical compositions of wine affects the variation of wine quality?

Data Introduction

-   The Wine Quality dataset was collected from Vinho Verde wine Samples, from the North of Portugal.
    -   The data was originally donated in 2009 by Professor Cortez
-   White Wine and Red Dataset were combined to create a dataset with 6,497 Observations
    -   The Variable good_wine was created to simplify quality (good wine>=7).
-   Response Variable: Good_wine
-   Categorical predictor Variable: Color of Wine
-   Quantitative predictor Variables: Physicochemical test results
    -   Fixed Acidity, Density, Alcohol content, pH, etc.
-   Goal: to produce a classification model that best explains how different chemical compositions of the Portuguese "Vinho Verde" wine affects the variation of the wine quality

Highlights from our Data Exploration
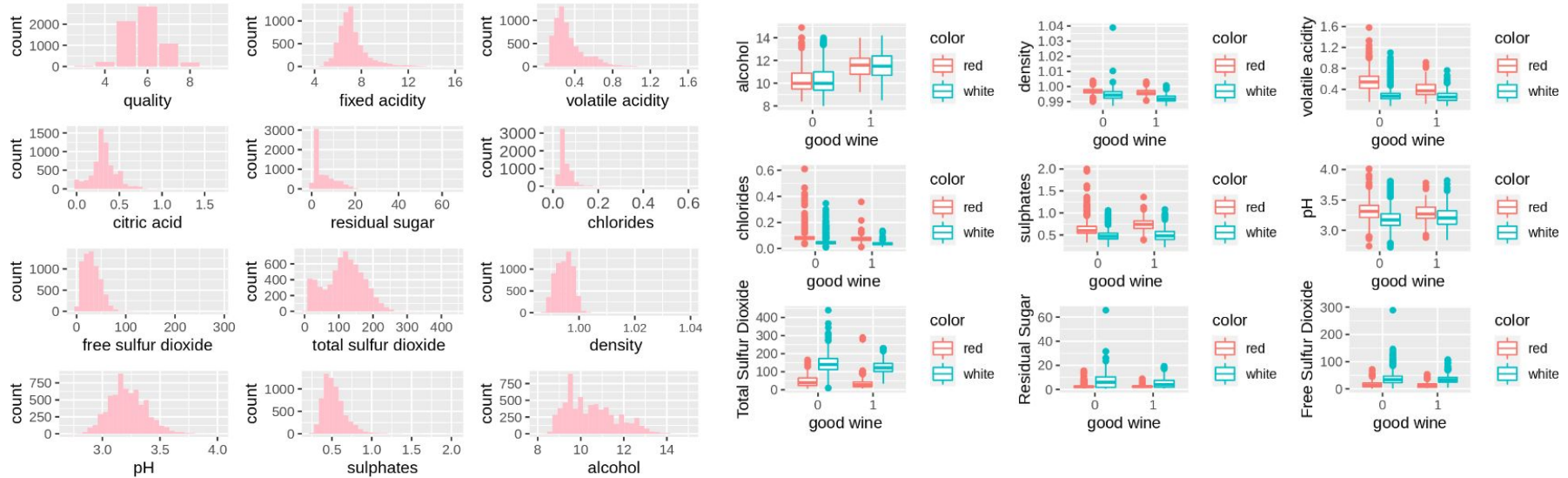
An interesting trend we found during our data analysis is that the variation between volatile acidity and total sulfur dioxide tended to be correlated with the quality of the wine (if a wine is good or not).

Wine Quality as seen through Logistic Regression

- We used stepwise AIC test to determine the best model without interactions.

- Our Final Model included all original variables except Citric Acid and Total sulfuric Acid.

- Our Model had an AIC score of 3808. 86 and an ROC-AUC estimate of 0.811.

- It performed similarly to a model with interaction terms based on the color of the wine.

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) : 407.674 + 0.472 \times fixed\ acidity - 3.397 \times volatile\ acidity + 0.214 \times residual\ sugar - 7.505 \times chlorides + 0.007 \times free\ sulfur\ dioxide - 427.514 \times density + 2.289 \times pH + 2.515 \times sulphates + 0.462 \times alcohol - 1.228 \times colorwhite$$

term	estimate	std.error	statistic	p.value
(Intercept)	407.674	74.178	5.496	0.000
c_fixed.acidity	0.472	0.075	6.304	0.000
c_volatile.acidity	-3.947	0.425	-9.289	0.000
c_residual.sugar	0.214	0.030	7.173	0.000
c_chlorides	-7.505	2.750	-2.729	0.006
c_free.sulfur.dioxide	0.007	0.003	2.561	0.010
c_density	-427.514	75.187	-5.686	0.000
c_pH	2.289	0.422	5.423	0.000
c_sulphates	2.515	0.324	7.755	0.000
c_alcohol	0.462	0.092	5.000	0.000
color_white	-1.228	0.242	-5.064	0.000

Some Interesting Findings From our Model

- Wine density, while very important, doesn't vary much, making it insignificant in our model.
- Citric Acid unexpectedly wasn't very important in predicting wine quality.
- Total Sulfur Dioxide was highly correlated with Free Sulfur dioxide, thus it was removed.
- Some Factors tended to have a negative effects on the quality of wine:
    - Chlorides: A major contributor to saltiness
    - Volatile Acidity: a kind of acid which makes wine tastes like vinegar.
- Some Factors Tended to have a positive effect on the quality on Wine:
    - pH: wine with low pH tastes tart and crisp, and higher pH adds to wine's ripeness
    - Sulphates:a food preservative widely used in winemaking which protects wine against oxidation, and maintains the flavor and freshness of wine.

Discussion and Conclusion

Key Findings/Conclusions:

- ***Our Findings point to the conclusion that under wine experts' standards, wine that tastes more salty and vinegary earns lower grades; on the other hand, wine taste riper, mellower, fresher receive higher grades.***

- ***pH and sulfate oxidation can be controlled using modern techniques and heavily influence wine quality***

Future Steps:

- Our research can be used as an impetus to fund and significantly improve further research in the process and industry of wine production
- Investigate data from more vineyards in different regions around the world to determine further trends
- Explore other tools for prediction: machine learning