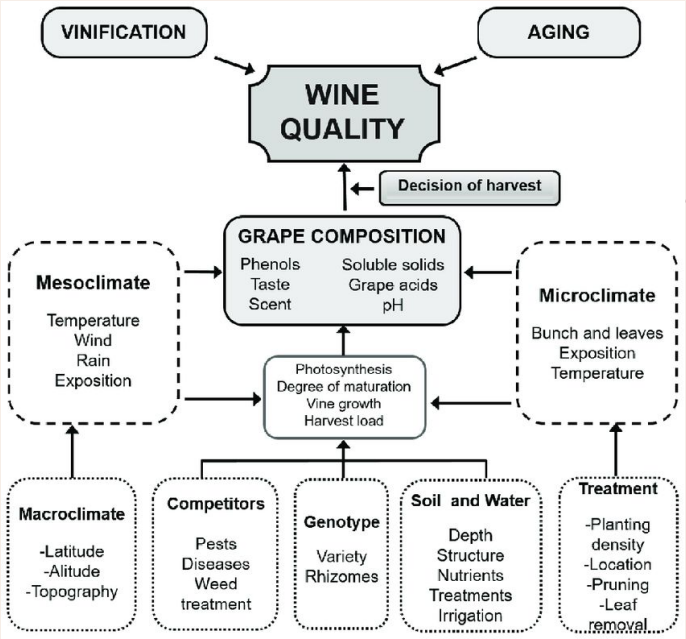
Red Wine Quality

Frankie Stillo, Yasmine Abdel-Rahman, Theo Prosise

Topic and Motivation

- Wine's importance to economy (1.3% GDP)
- The US alone consumes 4.3 billion bottles of wine annually
- Complexity of winemaking
- Variety of factors influence wine
- **Research question**: Which factors and properties can accurately predict the quality of wine?



Data

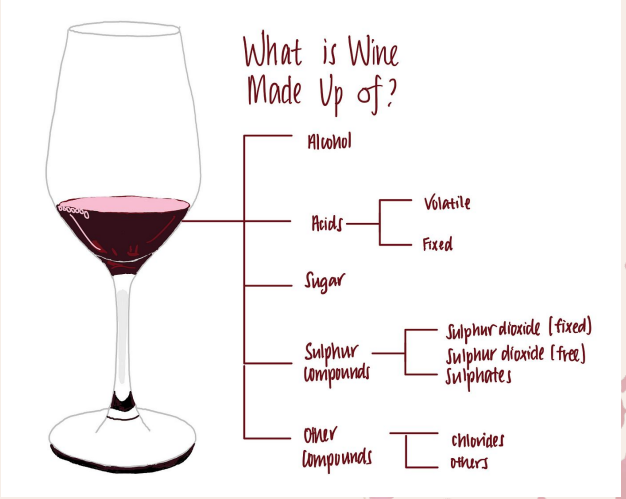
fixed.acidity	volatile.acidity	citric.acid	residual.sugar	chlorides	free.sulfur.dioxide	total.sulfur.dioxide	density	pH	sulphates	alcohol	quality
7.9	0.350	0.46	3.60	0.078	15	37	0.99730	3.35	0.86	12.8	8

variable	mean	sd	units	descriptions
fixed.acidity	8.320	1.741	g(tartaric acid)/dm^3	wine's natural acids
volatile.acidity	0.528	0.179	g(acetic acid)/dm^3	measure of the wine's gaseous acids that contributes to the smell and taste of vinegar in wine
citric.acid	0.271	0.195	g/dm^3	Boosts the acidity of wine during fermentation
residual.sugar	2.539	1.410	g/dm^3	natural grape sugars left in a wine after the alcoholic fermentation finishes.
chlorides	0.087	0.047	g(sodium chloride)/dm^3	adds to the saltiness of a wine
free.sulfur.dioxide	15.875	10.460	mg/dm^3	helps protect the wine from oxidation and spoilage
total.sulfur.dioxide	46.468	32.895	mg/dm^3	portion of SO_2 that is free in the wine plus the portion that is bound to other chemicals in the wine
density	0.997	0.002	g/cm^3	helps determine the alcohol content level of the final wine
pH	3.311	0.154	NA	can affect aroma, flavor, carbon dioxide absorption, tartrate precipitation, color, age-ability, fermentation rate, stability, and malolactic fermentation
sulphates	0.658	0.170	g(potassium sulphate)/dm^3	food preservative used to maintain the flavor and freshness of wine
alcohol	10.423	1.066	vol.%	Alcohol Content
quality	5.636	0.808	NA	Score given by experts

- Vinho Verde (Red) Wines
- 2004 - 2007
- most common physicochemical tests
- 12 numerical variables
- 1,599 observations
- Quality assessed by experts (0-10 scale)

Methodology

- RStudio analysis
- Regression Model Specification (continuous)
- 3 different linear regression models
- Interactive models
- Comparing adjusted R^2 and R^2



Models

Predictors	R^2	Adjusted R^2
Volatile Acidity and Alcohol Content	0.317	0.316
Sulphates, pH, Total Sulfur Dioxide, Alcohol Content	0.423	0.399
Every Chemical Variable	0.4789	0.380

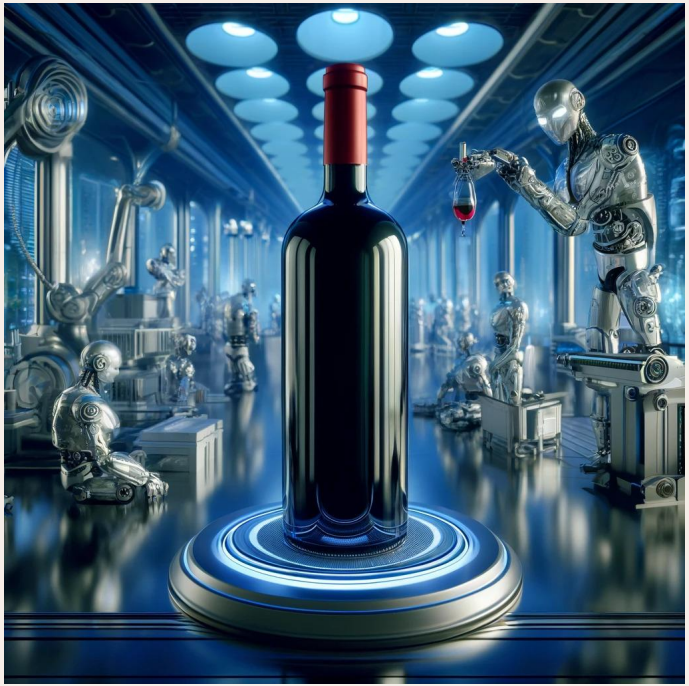
Results

- Model 2, Mix Model chosen
- Highest adjusted R^2
- Interactive model matches features
- Parallels conclusion of research

$$\begin{aligned}\hat{y} = &-696.7092 + 882.2127 \times \text{sulphates} + 211.4495 \times \text{pH} + 6.9559 \times \text{total sulfur} \\ &+ 68.9782 \times \text{alcohol} + 1445.558 \times \text{volatile acidity} \\ &+ 6533.450 \times \text{chlorides} - 268.4295 \times (\text{sulphates} \times \text{pH}) \\ &- 86.2912 \times (\text{sulphates} \times \text{alcohol}) - 20.7689 \times (\text{pH} \times \text{alcohol}) - \dots\end{aligned}$$

Conclusions / Future

- Modeling wine with many variables is difficult
- Interactive models better represent the data
- Future work with more combinations of variables
- White wine dataset can extrapolate conclusions
- Red wine vs white wine...
- Optimize wine production



Q & A

Thank you!