



WINE QUALITY DATASET ANALYSIS

What Makes Wine So Good?

KEY QUESTION AND PURPOSE

What chemical properties distinguish red and white wines, and how do these properties relate to wine quality scores?

Uncover actionable insights to improve wine quality and consistency.

WHO BENEFITS?

Wineries & producers (quality control, product development)

Distributors & retailers (product positioning)

Wine enthusiasts & researchers

DATA AND FEATURES

Data Source:

UCI Machine Learning Repository – Red & White Wine Quality datasets (CSV)

Unit of Analysis:

Each row = one wine sample

Features:

Fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free & total sulfur dioxide, density, pH, sulphates, alcohol, quality score

TOOLS AND PROCESS

Tools:

Jupyter Notebook

Python: pandas, numpy, matplotlib, seaborn

Process:

Data cleaning & preparation

Exploratory data analysis (EDA)

Visualization & summary statistics

SUMMARY STATISTICS

Alcohol: Red \approx 10.4%, White \approx 10.6%

pH: Red \approx 3.31, White \approx 3.20

Residual sugar: Red \approx 2.54 g/L, White \approx 6.84 g/L

	mean	median	std	min	max
fixed_acidity	7.215179	7.000000	1.319671	3.800000	15.900000
volatile_acidity	0.344130	0.300000	0.168248	0.080000	1.580000
citric_acid	0.318494	0.310000	0.147157	0.000000	1.660000
residual_sugar	5.048477	2.700000	4.500180	0.600000	65.800000
chlorides	0.056690	0.047000	0.036863	0.009000	0.611000
free_sulfur_dioxide	30.036654	28.000000	17.805045	1.000000	289.000000
total_sulfur_dioxide	114.109023	116.000000	56.774223	6.000000	440.000000
density	0.994535	0.994650	0.002966	0.987110	1.038980
ph	3.224664	3.210000	0.160379	2.720000	4.010000
sulphates	0.533357	0.510000	0.149743	0.220000	2.000000
alcohol	10.549241	10.400000	1.185933	8.000000	14.900000
quality	5.795677	6.000000	0.879772	3.000000	9.000000

VARIATION IN PROPERTIES

Most Variable:

Total sulfur dioxide, free sulfur dioxide, residual sugar

Least Variable:

Density, chlorides

	std
total_sulfur_dioxide	56.774223
free_sulfur_dioxide	17.805045
residual_sugar	4.500180
fixed_acidity	1.319671
alcohol	1.185933
quality	0.879772
volatile_acidity	0.168248
ph	0.160379
sulphates	0.149743
citric_acid	0.147157
chlorides	0.036863
density	0.002966

Most variable: total_sulfur_dioxide (std=56.77)

Least variable: density (std=0.0030)

RED VS. WHITE DIFFERENCES

Red Wines: Higher acidity, higher sulphates

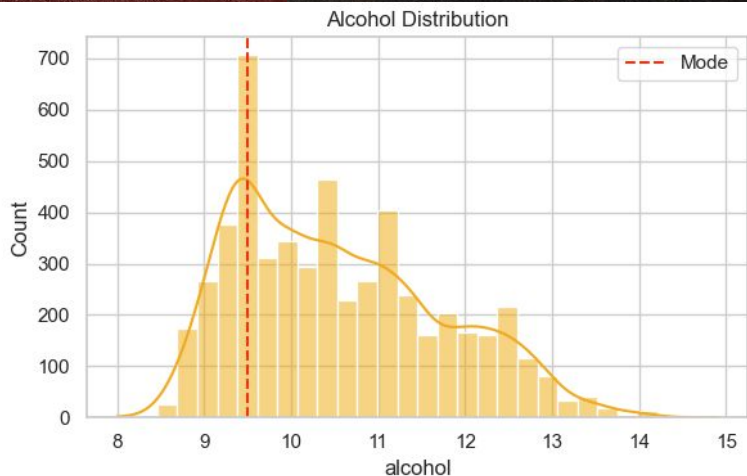
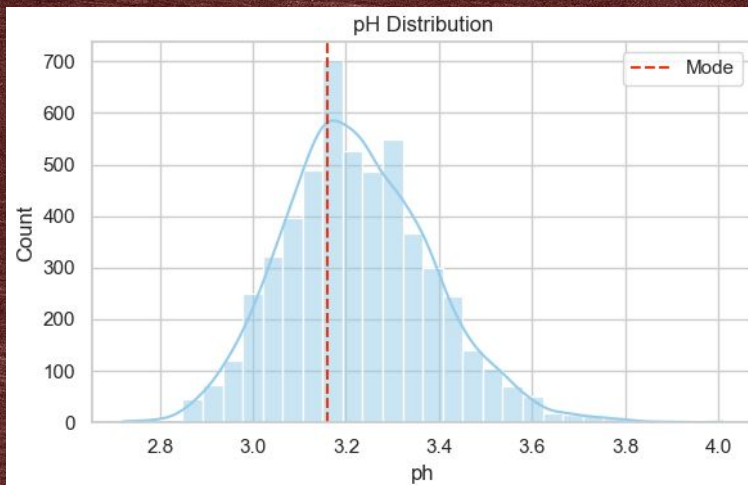
White Wines: Higher residual sugar, higher sulfur dioxide

	fixed_acidity			residual_sugar			free_sulfur_dioxide			total_sulfur_dioxide			
	mean	std	median	mean	std	median	mean	std	median	mean	std	median	mean
type													
red	8.310596	1.736990	7.900000	0.529478	1.352314	2.200000	0.088124	10.447270	14.000000	46.825975	33.408946	38.000000	0.996709
white	6.839346	0.866860	6.800000	0.280538	4.861646	4.700000	0.045905	17.210021	33.000000	137.193512	43.129065	133.000000	0.993790

MOST COMMON VALUES

pH Mode: Red 3.3, White 3.1

Alcohol Mode: 9.5% (both)

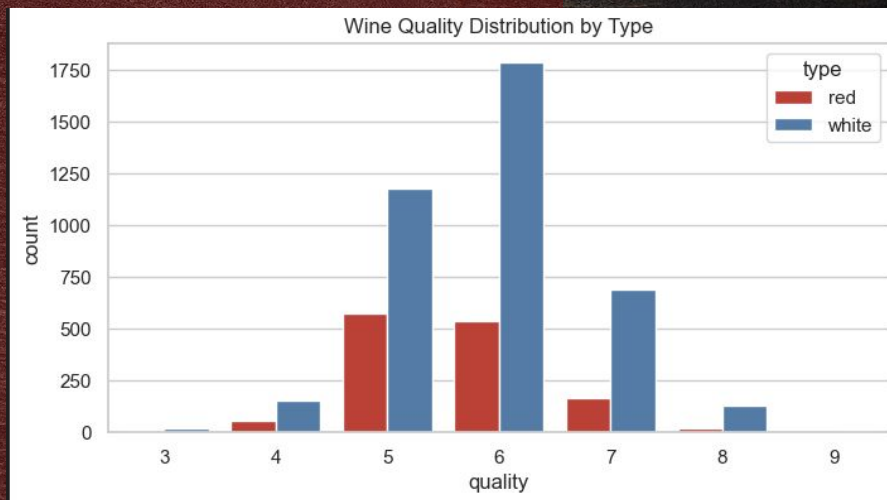


QUALITY DISTRIBUTION

Most Common Score: 6 (both types)

Red: Clustered at 5 and 6

White: More samples at 6 and 7

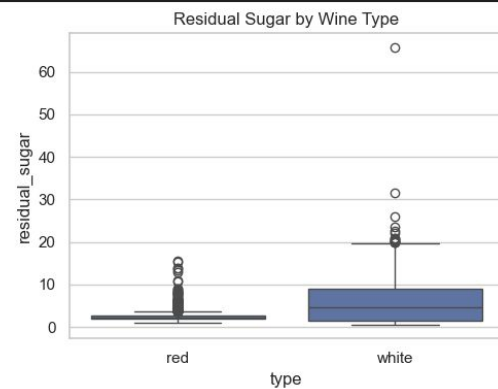
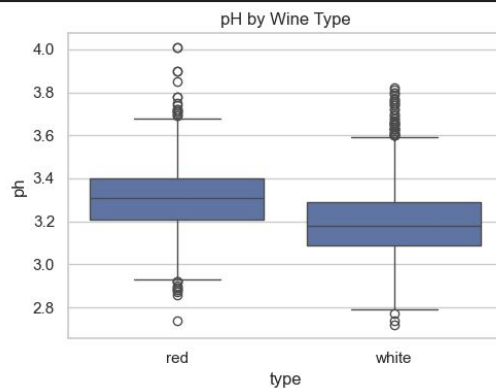
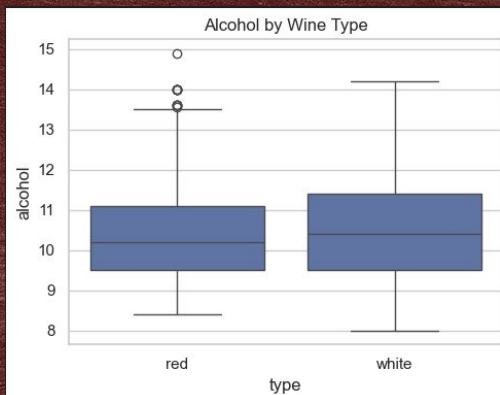
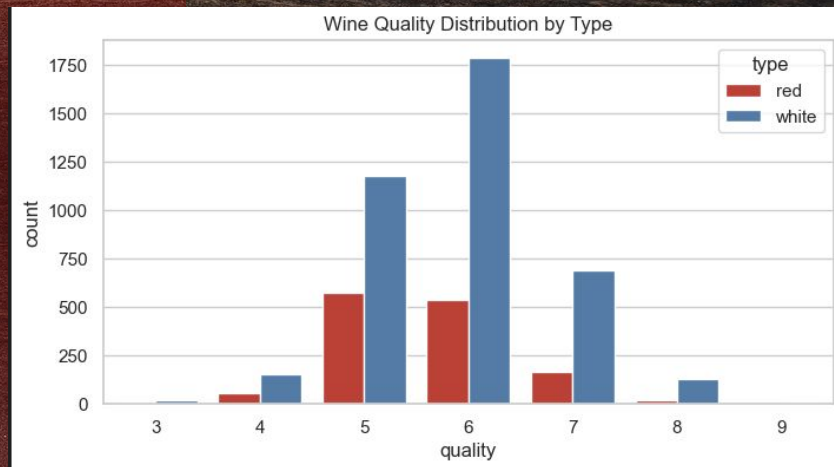


KEY PLOTS

Conclusion:

Both types cluster around quality 6

Alcohol content is a key differentiator



RECOMMENDATIONS

For Producers:

Control sugar and sulfur dioxide (especially in whites)

Monitor density & chlorides as baselines

For Marketing:

Highlight wines with pH \approx 3.3, alcohol \approx 10.5%

Promote wines with quality scores of 6+

FUTURE WORK

Predictive modeling to forecast wine quality

Identify key quality drivers

Integrate sensory/tasting data

Analyze trends over time (if vintage data available)

APPENDIX & Q&A

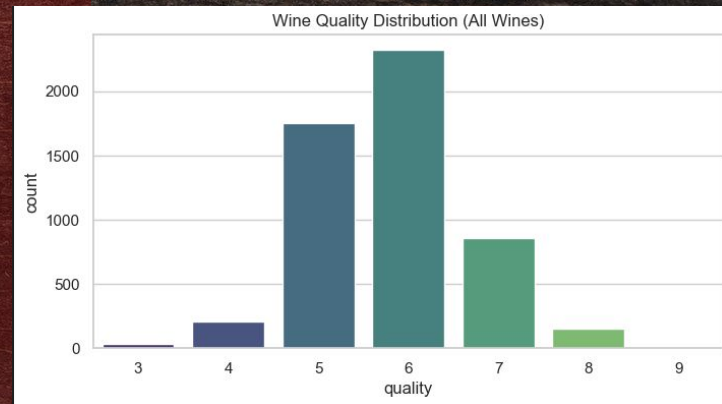
Most wines are rated between 5 and 7.

Quality scores are not evenly distributed

Score of 6 is the most common quality rating in the dataset.

Indicates that the majority of wines are of average to slightly above-average quality.

Useful reference for producers to understand how their wines compare to the overall market.





THANK YOU