Key Findings from the Wine Quality Research

The WinequalityEDA notebook provides an exploratory data analysis of the physicochemical properties influencing the quality of red wine. Below are the main insights drawn from the presented data:[1]

1. Descriptive Statistics

- The wine dataset consists of 1,599 samples and 12 attributes, which include chemical features like acidity levels, sugar, sulfur dioxide content, density, pH, sulphates, alcohol percentage, and a quality score.
- **Quality scores** range from 3 to 8, with a mean value of 5.64 and a median of 6, indicating a predominance of average-quality wines.
- **Alcohol content** has a mean of 10.42% and ranges from 8.4% to 14.9%. **Fixed acidity** ranges from 4.6 to 15.9, and volatile acidity ranges from 0.12 to 1.58.

2. Correlations with Quality

- **Alcohol content** has the strongest positive correlation with wine quality ($ r = 0.480 $), confirming that higher alcohol wines tend to score better in quality assessments.
- **Sulphates** and **citric acid** also show positive associations ($ r = 0.249 $ and $ r = 0.228 $, respectively), suggesting these chemicals may enhance perceived quality.
- **Volatile acidity** has a negative correlation with wine quality ($ r = -0.395 $), indicating that higher levels reduce quality.
- Other negative correlates include **density** ($ r = -0.184), $ ** chlorides ** (r = -0.131), and ** totalsulfurdioxide ** (r = -0.178 $), implying that wines with higher density or sulfur/chloride content are generally rated lower.

3. Distributions and Summary

- Most features display typical measurement ranges for red wine, without major outliers or anomalies.
- The dataset contains only a modest spread of quality scores, with most wines considered “average.”
- The chemical attributes are within plausible ranges:
    - **pH**: Mean 3.31 (slightly acidic to neutral).

- **Sulphates**: Mean 0.66, positively linked to quality.
- **Free sulfur dioxide**: Mean 15.87, total sulfur dioxide mean 46.47.

4. Statistically Significant Relationships

- The correlation matrix suggests that, for predictive modeling, **alcohol**, **volatile acidity**, and **sulphates** are the best predictors of wine quality.
- **Fixed acidity** and **citric acid** are related to each other ($ r = 0.67 $) but weakly linked to quality.

Conclusion:

The main findings of the analysis are that **higher alcohol content, sulphates, and citric acid are associated with better wine quality**, whereas **high volatile acidity, density, chlorides, and sulfur dioxide relate to lower quality**. Most wines in the dataset have middling quality scores, and the chemical properties align with general expectations for red wine.[1]

⁂

1. WinequalityEDA.ipynb