

# WINE QUALITY CASE PROJECT

**Analyze the chemical properties of red wines to identify key factors influencing quality, helping winemakers optimize production and improve product appeal.**

**A case study carried out by Vansh Sinha**

# OUTLINE

- Introduction
- Data cleaning
- Key visualizations
- Decision Tree  
+  
Random Forest



# INTRODUCTION

## SUMMARY OF DATASET

This dataset contains various physicochemical properties and quality ratings for red wine samples. The features (fixed acidity, volatile acidity, citric acid, etc.) are used to predict the quality of the wine, which is rated on a scale from 0 to 10.

## COLUMN DESCRIPTIONS

- **fixed acidity:** Non-evaporating acids.
- **volatile acidity:** Evaporating acids.
- **citric acid:** Adds freshness.
- **residual sugar:** Remaining sugar.
- **chlorides:** Salt content.
- **free sulfur dioxide:** Unbound SO<sub>2</sub>.
- **total sulfur dioxide:** Total SO<sub>2</sub>.
- **density:** Mass per volume.
- **pH:** Acidity level.
- **sulphates:** Prevents spoilage.
- **alcohol:** Alcohol content.
- **quality:** Quality score.

# DATA CLEANING

Resolving issues of messy and dirty data



# ISSUES WITH DATASET



DIRTY DATA

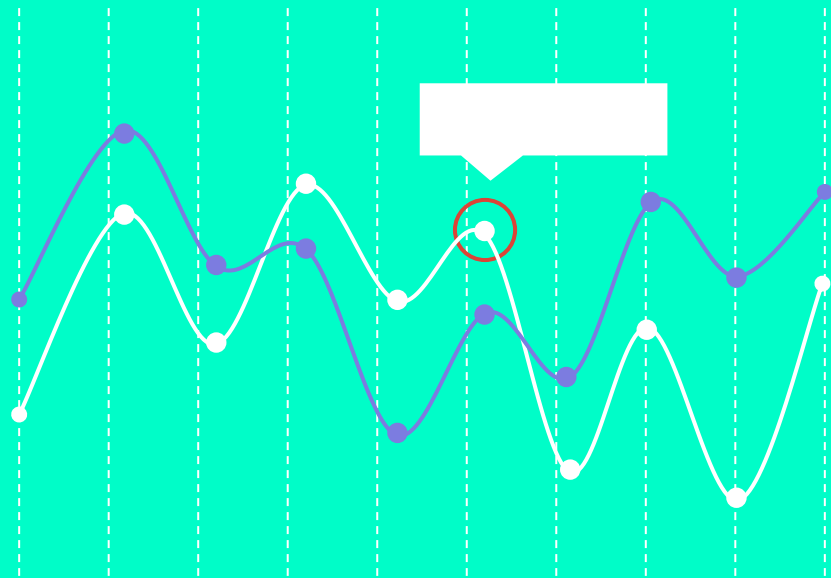
MESSY DATA

- **Duplicate Rows**

**NOT APPLICABLE**

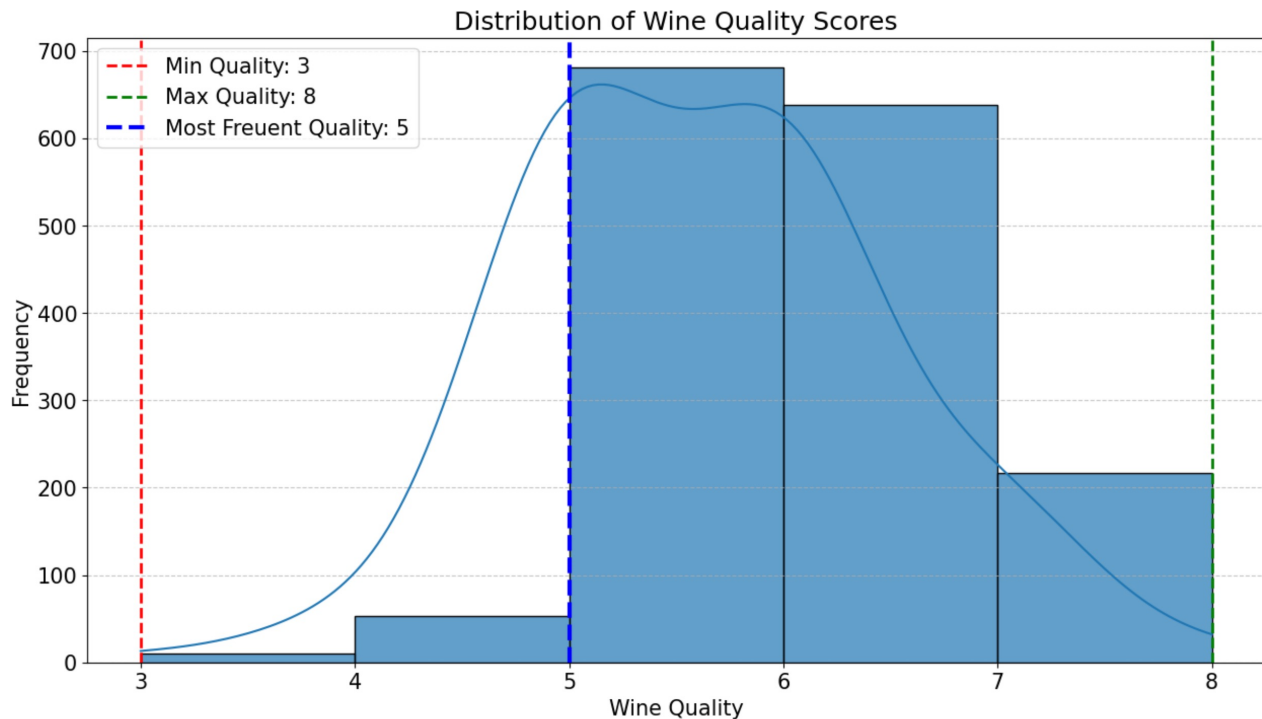
# KEY VISUALIZATIONS

Visualizations from the dataset



## CHART 1: WINE QUALITY SCORES

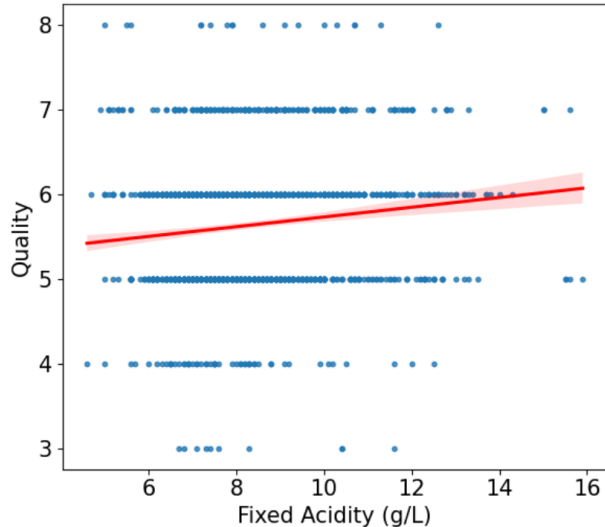
Most wines have a quality score of 5; quality ranges from 3 (min) to 8 (max), showing a slight right skew.



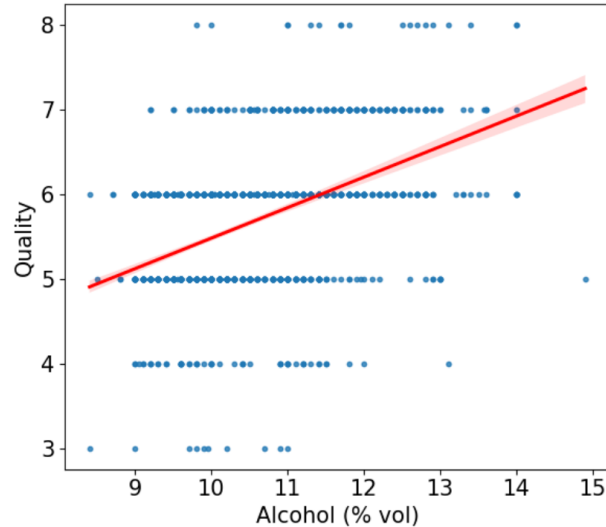
## CHART 2: CORRELATION OF METRICS WITH QUALITY

- **Fixed Acidity**: Weak positive effect on quality (0.12).
- **Alcohol**: Moderate positive effect on quality (0.48).
- **Free Sulphur Dioxide**: Negligible effect on quality (-0.05).

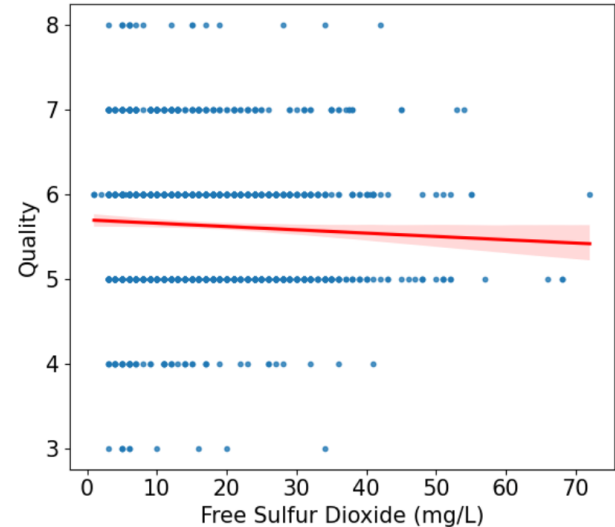
Fixed Acidity vs. Quality  
Correlation: 0.12



Alcohol vs. Quality  
Correlation: 0.48



Free Sulfur Dioxide vs. Quality  
Correlation: -0.05





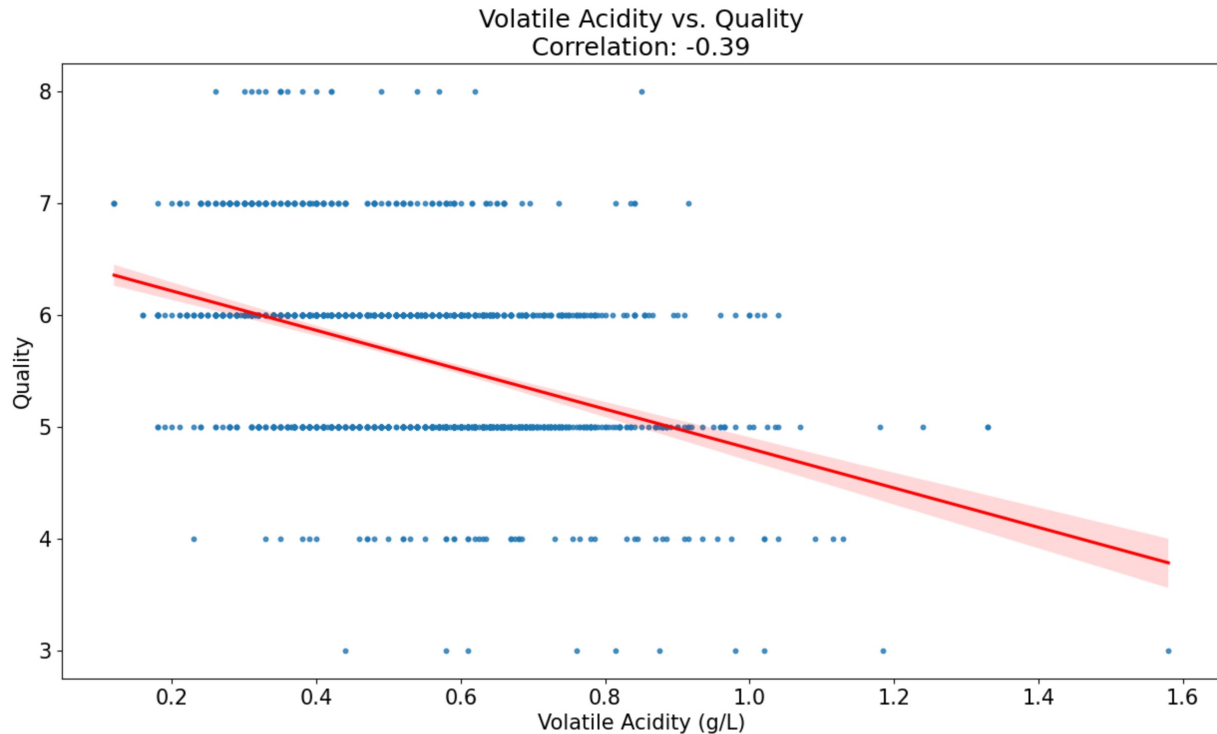
### CHART 3: AVG. RESIDUAL SUGAR VS QUALITY

Residual sugar is almost the same for both low and high-quality wines, with only a -2.17% difference.



## CHART 4: VOLATILE ACIDITY VS QUALITY

Higher volatile acidity moderately correlates with lower wine quality (-0.39).



# DECISION TREE + RANDOM FOREST



# WORKFLOW

- Train-Test Split
- Decision Tree
- Random Forest (n\_estimators=100)

DECISION TREE

ACCURACY

50.7%

VS

RANDOM FOREST

ACCURACY

66.9%



THANK YOU!