Finlatics
ENABLING INSIGHTS

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 Aryan Singhal

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 SO2.
- **total sulfur dioxide:** Total SO2.
- **density:** Mass per volume.
- **pH:** Acidity level.
- **sulphates:** Prevents spoilage.
- **alcohol:** Alcohol content.
- **quality:** Quality sore.

Data cleaning

Resolving issues of messy and dirty data

Issues with Dataset

Dirty Data

- Duplicate Rows

Messy Data

NOT APPLICABLE

Key visualizations

Visualizations from the dataset

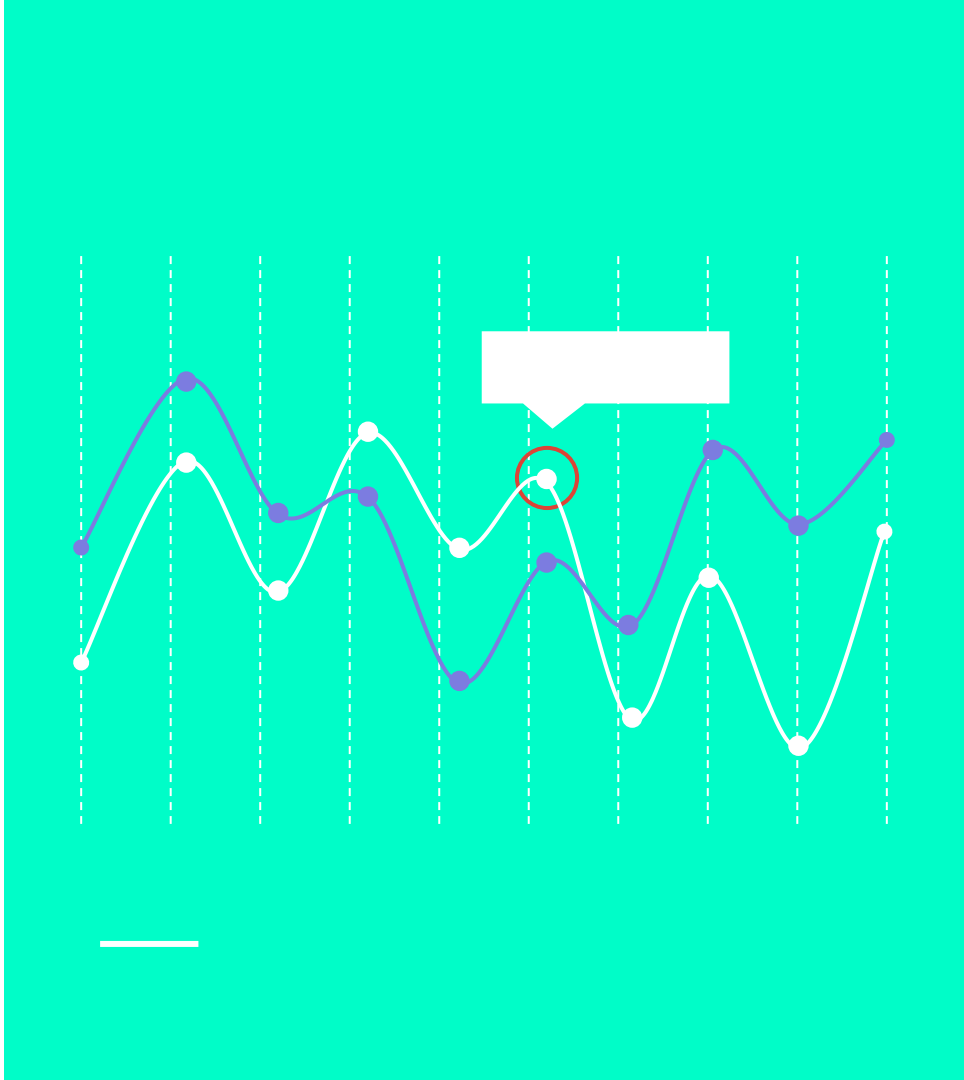


Chart 1: Wine Quality Scores

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

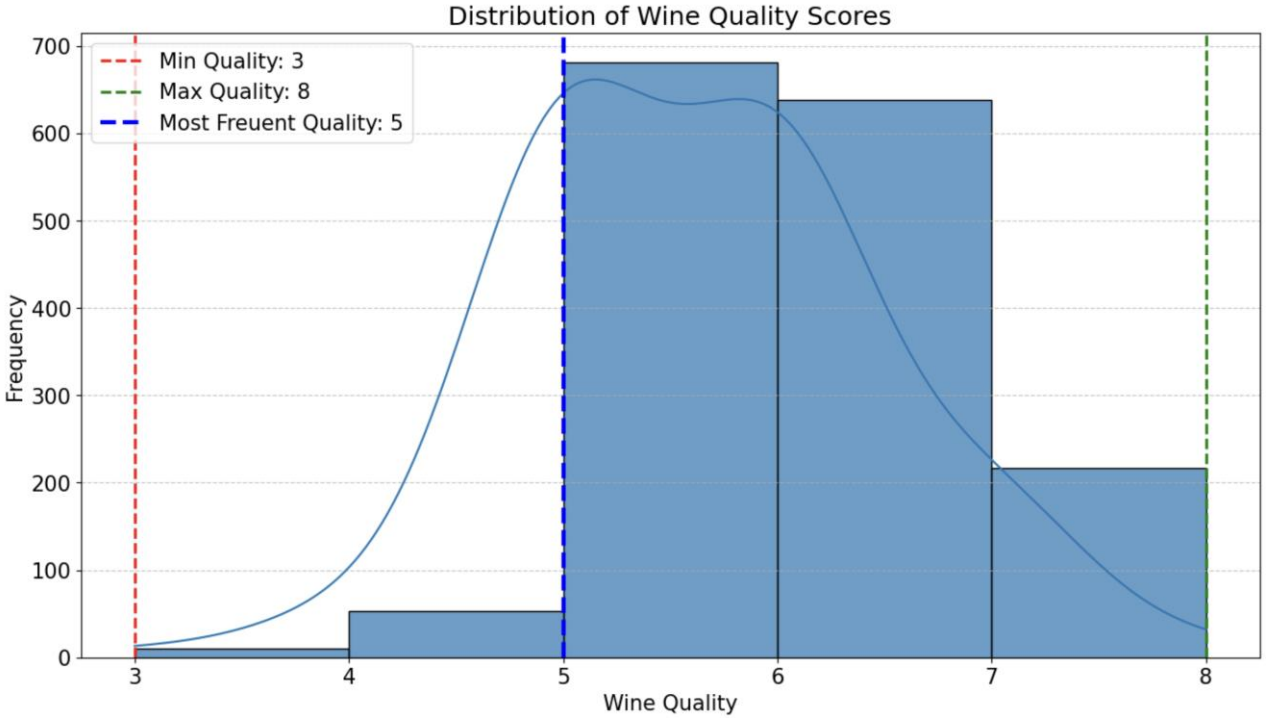
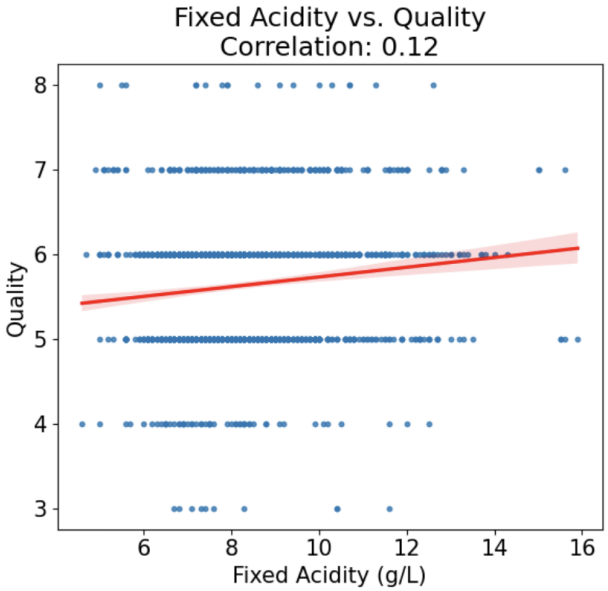
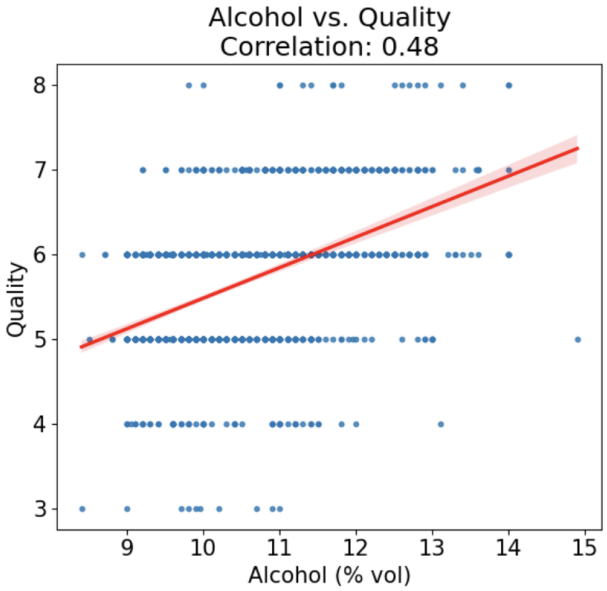


Chart 2: Correlation of metrics with quality

- ***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).





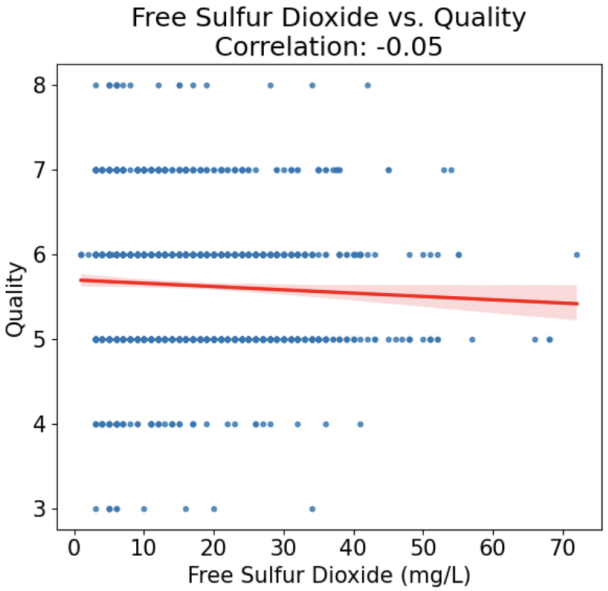
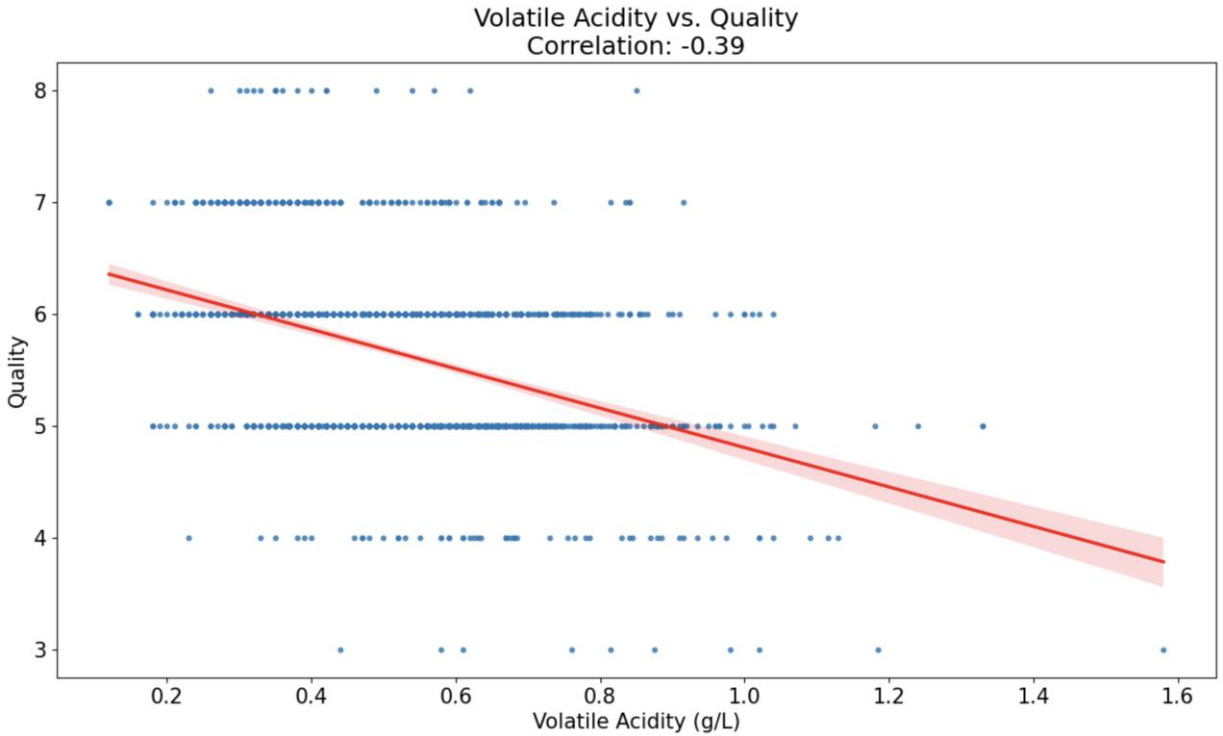


Chart 3: Avg. Residual sugar vs Quality

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

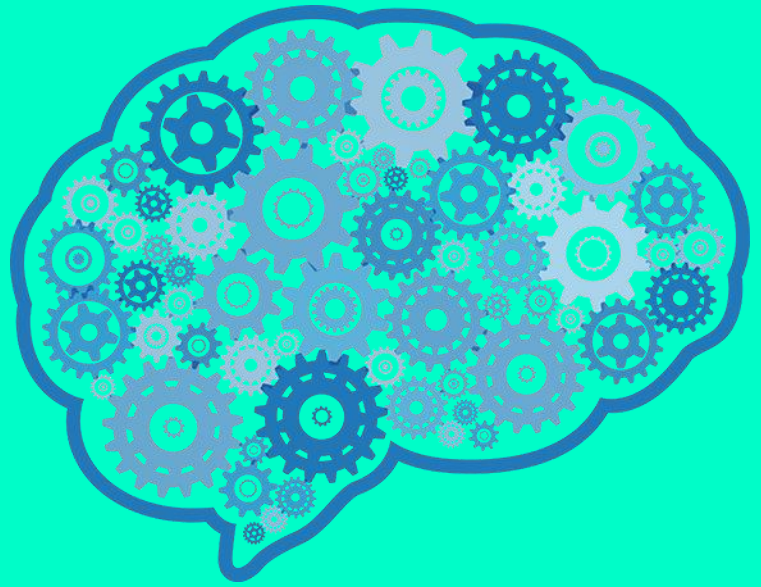
Higher volatile acidity moderately correlates with lower wine quality (-0.39).



Decision Tree

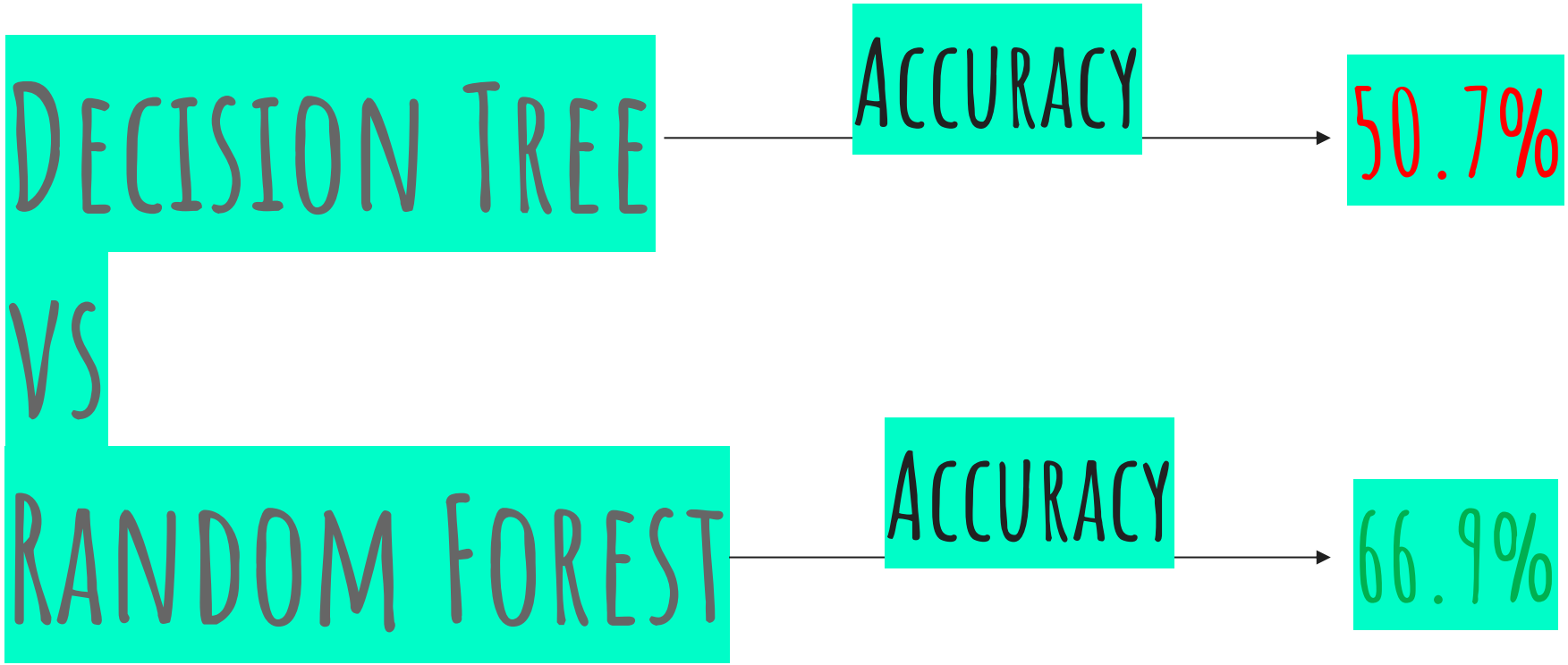
+

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!