Vineyard Vintage Winery

Project By Ray Onsongo

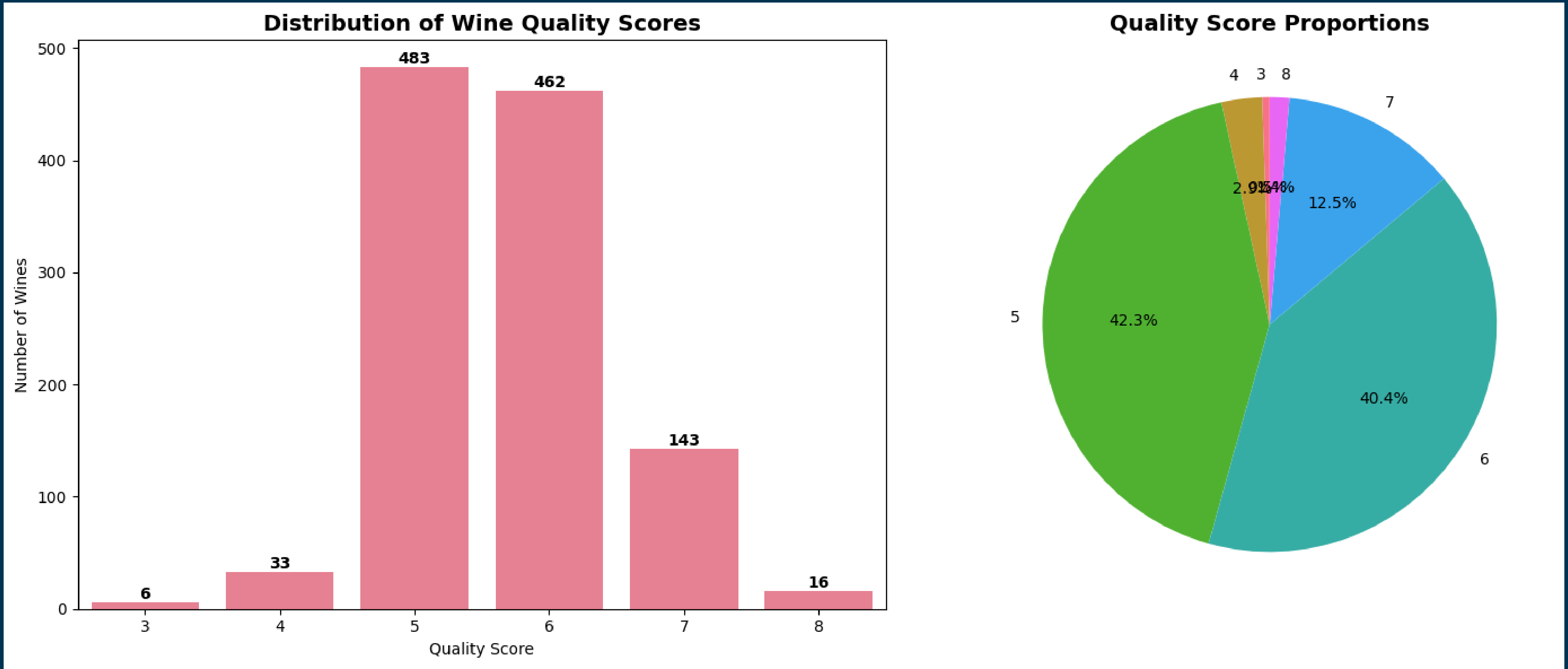
INTRODUCTION

Subtitle

Business Understanding Section

- Business Problem: Wine Quality Classification
- **Stakeholder:** Quality Assurance Manager, Vintage Vineyards Winery
- **Business Context:** Our winery has been experiencing inconsistent quality across wine batches, leading to:
- Price reductions for lower-quality wines
- Customer dissatisfaction and returns
- Inefficient allocation to premium vs. standard distribution channels
- **Business Objective:** Develop a predictive model that can classify wine batches as "Premium Quality" (quality ≥ 7) or "Standard Quality" (quality < 7) based on measurable chemical properties.
- **Business Value:**
- Route premium wines to high-margin channels ($25+ bottles)
- Identify underperforming batches early for blending or correction
- Optimize production parameters to increase premium yield
- Estimated potential revenue increase: 15-20% through better quality control """

Exploratory Data Analysis



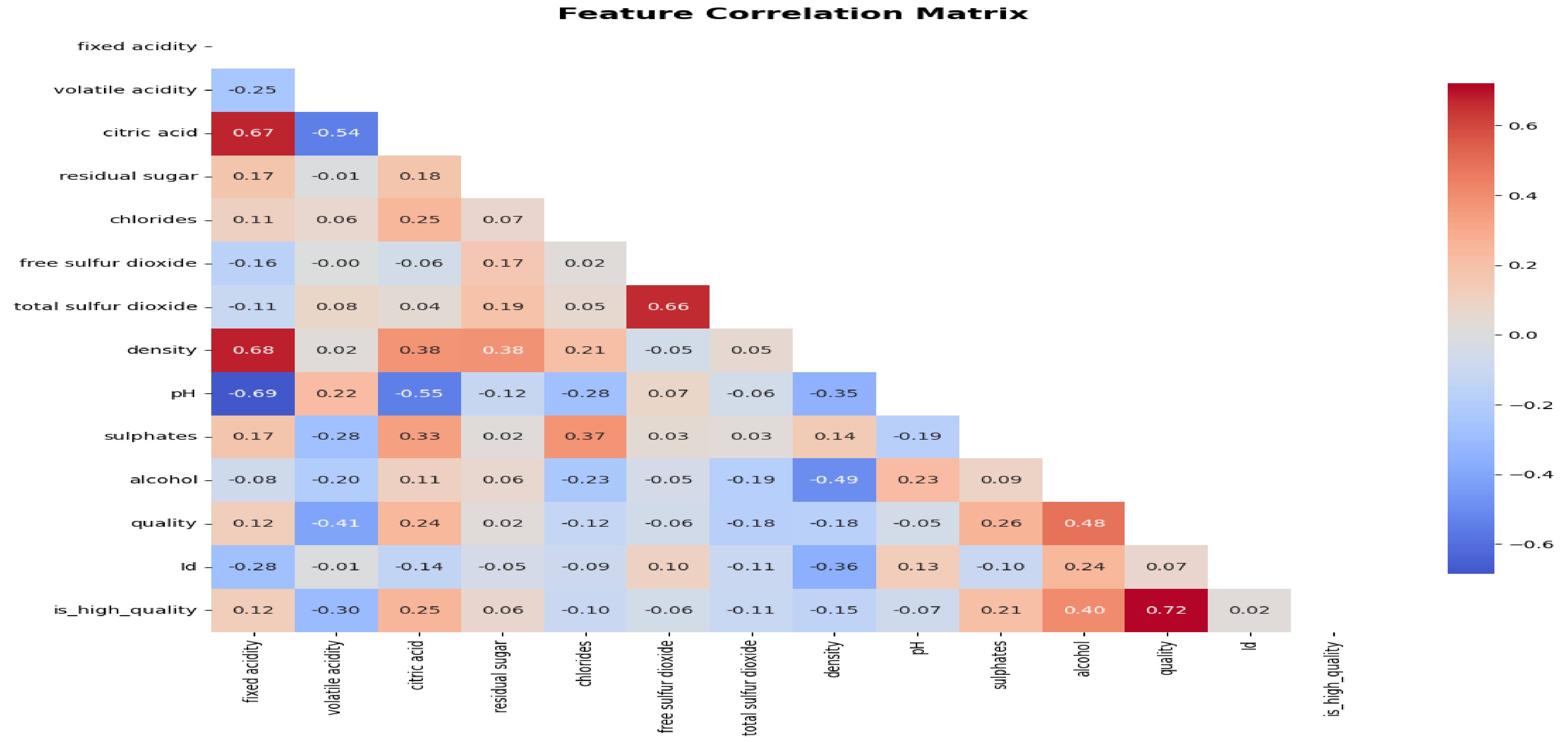
Final Threshold Decision

- STEP 5: FINAL THRESHOLD DECISION ===
- Selected threshold: 7+ = High Quality
- Rationale: This creates a meaningful business distinction between premium and standard wines

- Final class distribution:
- • Standard Quality (0): 86.1%
- • High Quality (1): 13.9%

- Dataset ready for further EDA and modeling!
- New columns added: 'is_high_quality' (binary) and 'quality_category' (string)

Correlation Analysis



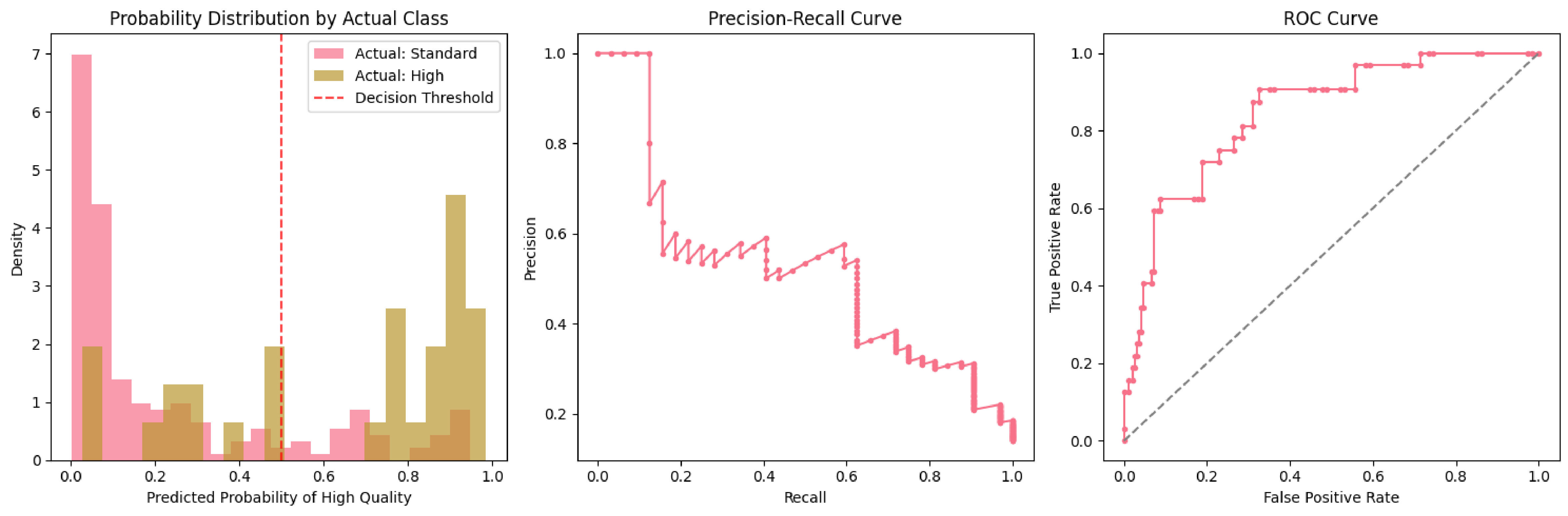
Baseline Modeling & Preprocessing

- 1. TRAINING vs TESTING PERFORMANCE

Metric	Training	Testing	Difference
Accuracy	0.8316	0.7860	-0.0456
Recall	0.8590	0.6250	-0.2340
Precision	0.8145	0.3509	-0.4636
F1-Score	0.8361	0.4494	-0.3867

- MODEL INTERPRETABILITY - FEATURE COEFFICIENTS
- Top 5 most influential features:
  - alcohol : +1.2210 (increases high quality probability)
  - sulphates : +0.7233 (increases high quality probability)
  - citric acid : +0.6287 (increases high quality probability)
  - total sulfur dioxide: -0.5702 (decreases high quality probability)
  - volatile acidity : -0.5137 (decreases high quality probability)

Prediction Probabilities



Baseline Model Performance

- Key Observations:
- Significant Overfitting: Large gap between training and test performance, especially for recall (-0.234) and precision (-0.4636)
- Good Recall, Poor Precision: Model finds 62.5% of high-quality wines but has many false positives (only 35% precision)
- Reasonable Overall Accuracy: 78.6% on test set

- Business Implications:
- The Good: Your model successfully identifies 62.5% of high-quality wines (good recall)
- The Problem: When it predicts "high quality," it's only correct 35% of the time (poor precision). This means:
- 65% of wines labeled as "premium" are actually standard quality
- This could damage brand reputation and customer trust

Iterative Modeling & Hyperparameter Tuning

- COMPREHENSIVE EVALUATION: DT Conservative
- TRAINING vs TESTING PERFORMANCE

Metric	Training	Testing	Difference
Accuracy	0.9238	0.8472	-0.0766
Recall	0.9822	0.7188	-0.2635
Precision	0.8794	0.4694	-0.4100
F1-Score	0.9280	0.5679	-0.3601

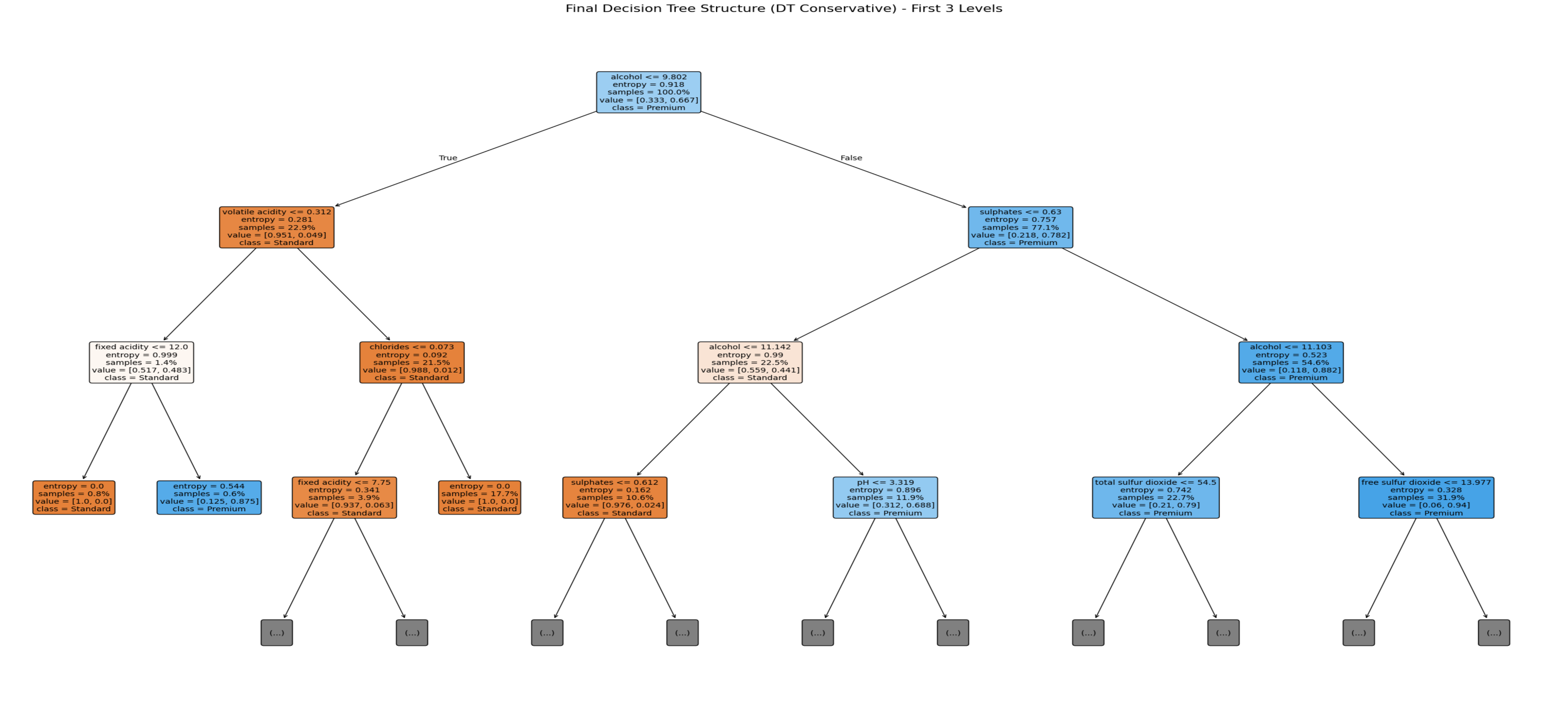
- BUSINESS-FOCUSED METRICS
  - Recall (Sensitivity) : 0.719 - Ability to find High Quality wines
  - Precision : 0.469 - Accuracy when predicting High Quality
  - Specificity : 0.868 - Ability to identify Standard Quality correctly
  - False Positive Rate : 0.132 - Standard wines mislabeled as High Quality
  - False Negative Rate : 0.281 - High Quality wines missed

FINAL MODEL SELECTION & BUSINESS JUSTIFICATION

- DT Conservative
- BUSINESS JUSTIFICATION:
- • Highest Recall (71.9%): Finds the most high-quality wines
- • Good Balance: Best F1-score among all models
- • Acceptable Precision: 47% precision is reasonable for initial screening
- • Business Impact: Identifies 72% of premium wines vs 59% in baseline

- BUSINESS IMPACT ANALYSIS:
- • High-Quality Wines Correctly Identified: 23 out of 32 (71.9%)
- • Standard Wines Incorrectly Labeled Premium: 26 out of 197 (13.2%)
- • Missed High-Quality Opportunities: 9 wines

Decsision Tree Conservative



Final Project Summary

- PROJECT OUTCOMES: ✅ Built and evaluated multiple Decision Tree models ✅ Achieved 72% recall (finding high-quality wines) ✅ Improved precision from 36% to 47% over baseline ✅ Reduced false premium rate from 17% to 13% ✅ Identified key quality factors: alcohol, sulphates, acidity
- 🎯 BUSINESS VALUE: • The model can identify 72% of premium wines automatically • Reduces manual tasting workload by effective pre-screening • Provides consistent, data-driven quality assessment • Helps optimize production parameters for better quality
- 🔧 TECHNICAL ACHIEVEMENTS: • Best Model: DT Conservative • Key Metric: Recall = 71.9% • Model Discriminatory Power: AUC = 0.788 • Feature Selection: Used 11 chemical properties
- 🌟 FINAL RECOMMENDATION: Deploy the DT Conservative model as a pre-screening tool in your quality control process, with expert tasting as the final validation step.

END
Thank You