
In Vino Veritas
GA Data Science Final Project
3/7/2016
Wen Lu

How to discern wine quality?

By description: *Balance, Length, Depth, Complexity, Finish*

By measurements: *Physiochemical characters*

Wine making is considered an art.

- Is there a formula for a quality wine?

- What basic properties are the formula for a good wine?

- Do white wine and red wine share the same formula?

- Are there any characters that distinguish white wine and red wine?

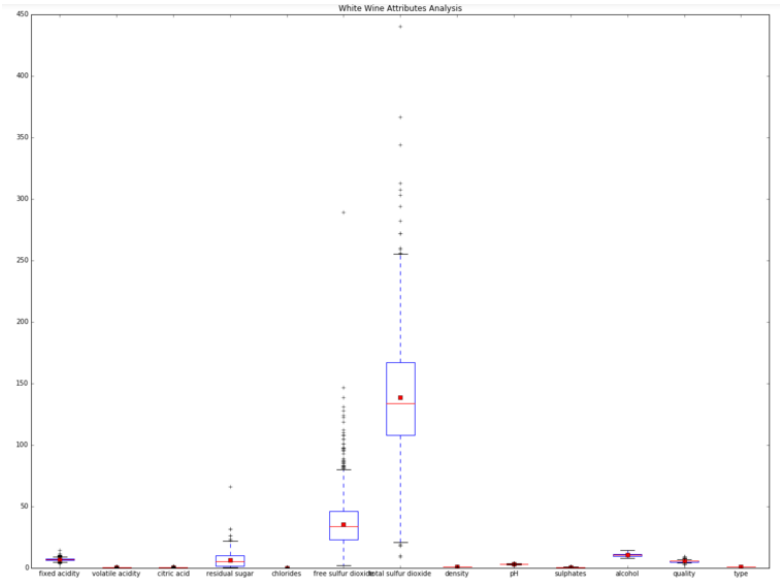
Data Source: UCI Machine Learning Repository

Wine Quality Data Set

http://mlr.cs.umass.edu/ml/datasets/Wine+Quality

Two datasets: White Wine & Red Wine

- 12 attributes
- no missing values
- sample size >1k (1599 vs 4898)
- some noises
- different measurement units
- some features correlated



	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.0	0.27	0.36	20.7	0.045	45	170	1.0010	3.00	0.45	8.8	6
1	6.3	0.30	0.34	1.6	0.049	14	132	0.9940	3.30	0.49	9.5	6
2	8.1	0.28	0.40	6.9	0.050	30	97	0.9951	3.26	0.44	10.1	6
3	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.40	9.9	6
4	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.40	9.9	6

Framework

- **EDA**
  - Feature distribution graph (to be done)
  - Standardization
  - Benchmark (Dummy Classifier)

- **Supervised Learning**
  - Logistic Regression
  - Lasso
  - Random Forest (importance score)
  - SVM

- **Unsupervised Learning**
  - [Concatenate datasets]
  - PCA
  - K-means Clustering
  - Visualization

- **Challenges**
  - Perform on 3 datasets
  - How to determine good white wine formula and good red wine formula are drastically different or not?