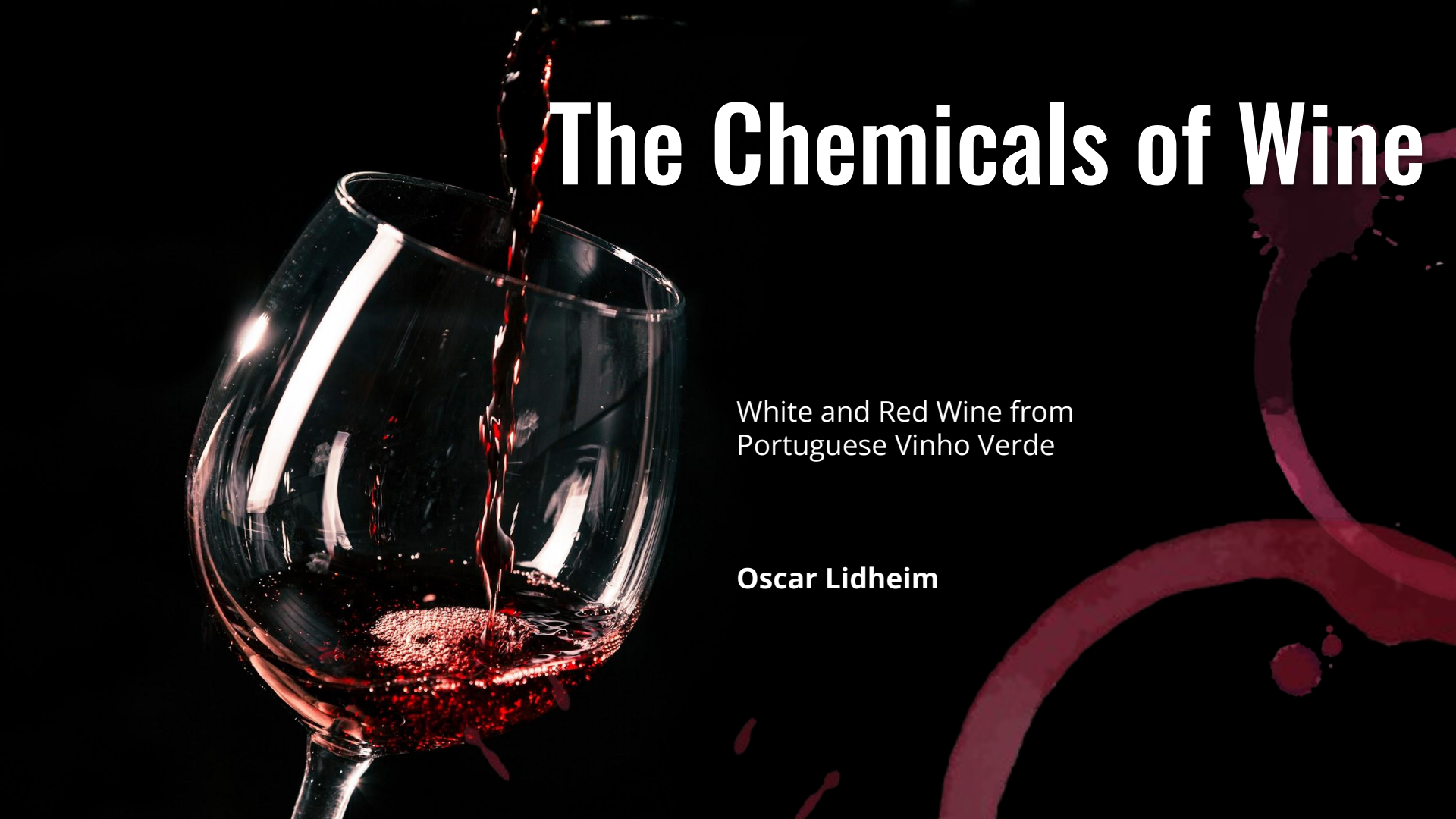


The Chemicals of Wine

White and Red Wine from
Portuguese Vinho Verde

Oscar Lidheim

AGENDA

01. Dataset and Wine

02. Clusters

03. PCA

04. Conclusion

Dataset

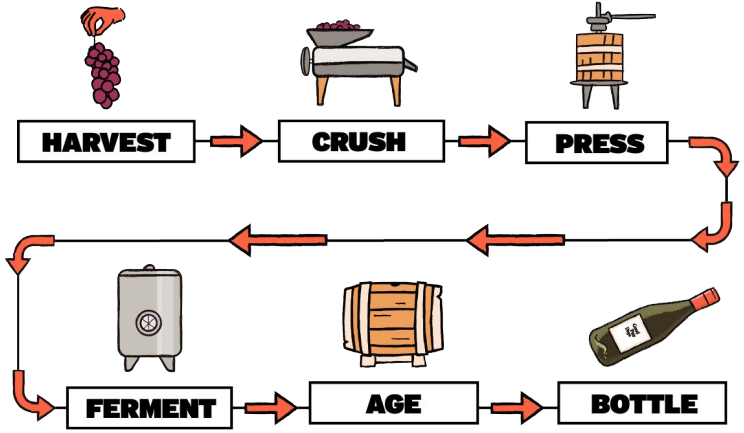
Vinho Verde in Portugal

Data used in several research studies

6,600 wines, objectively measured (chemical tests) on 11 features and 1 expert rating on quality

Data:
1.6k red
4.9k white

Dimensions
6.6k x 12



Wine Characteristics

1

Chemical Composition:

Fixed Acidity (To low flat, too high sour)
Volatile Acidity (keep low for good taste)
Citric Acid (Organic acid, citrusy flavour)
pH (acidity measurement, preference)
Residual Sugar (Sweetness / Dryness)
Chlorides (Mid level, too high salty)
Density (No taste - fullness)

2

Sulfur Dioxide:

Free Sulfur Dioxide
Total Sulfur Dioxide
(added preservative)

3

Taste and Quality:

Sulphates (results yeast metabolism, minimal taste)
Alcohol (high taste, richness)

4

Quality:

Rating 0-10

Median of at least 3 expert opinions

Red and White

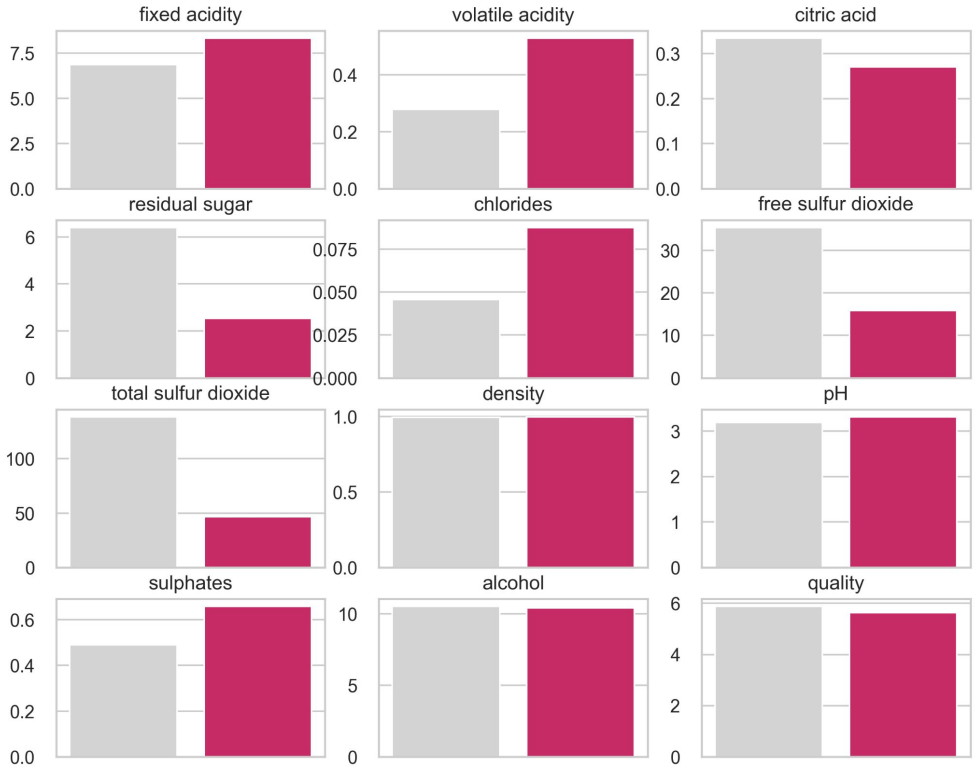
Distinct Features (mean values):

White:

- Higher sugar (3x)
- Higher sulfur dioxide
- Higher citric acid (PH)

Red:

- Higher volatile acidity
- Higher density



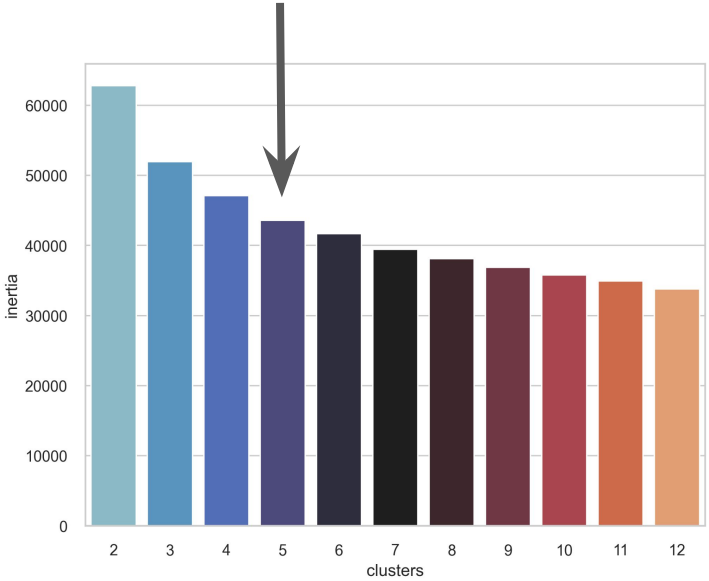
KMeans

5 Clusters (elbow)

Perfect clusters by red/white wine

Clusters:

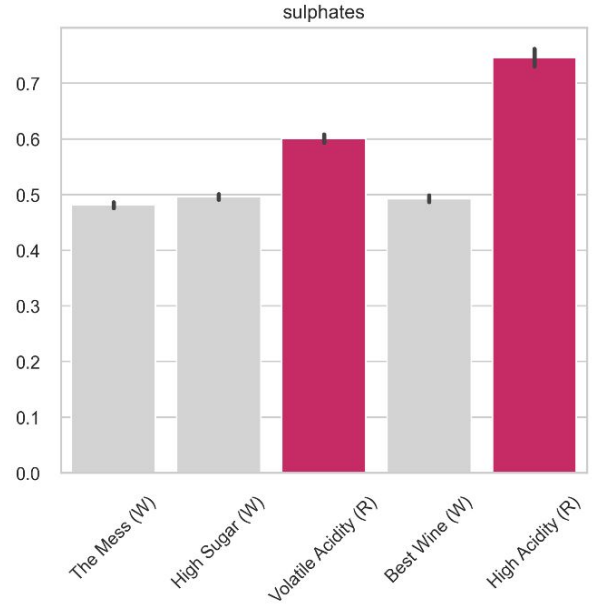
1) The mess (1600) -> RANDOM BLEND
2) High Sugar (1600) -> FOR THE SWEET TOOTH
3) High Volatile Acidity (1000) -> BAD TASTE
4) Best Wine (1700) -> WELL BALANCED MID RANGE MOST VALUES
5) Fixed Acidity + Citric Acid (600) -> GOOD RED

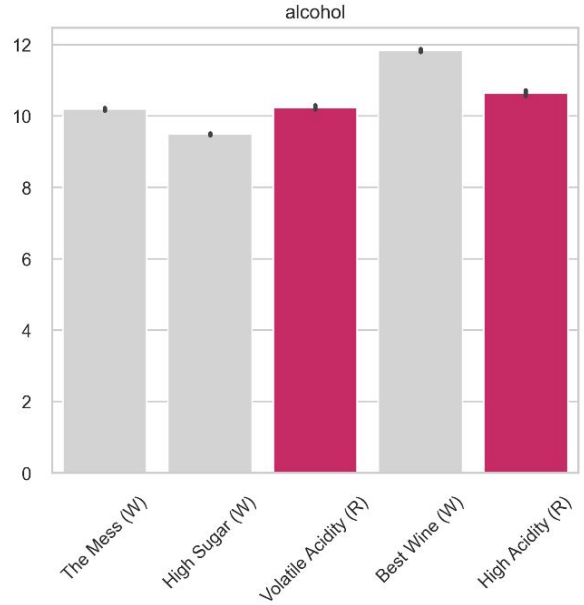


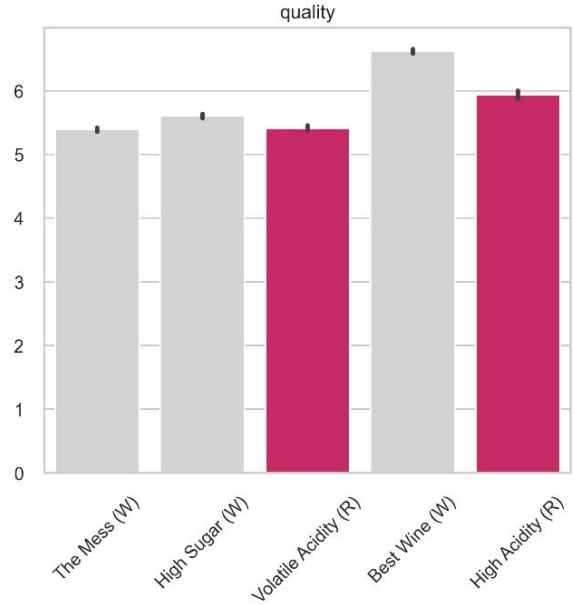
Quality by Cluster

Alcohol and Quality highest correlation

White wine seems to be this regions speciality



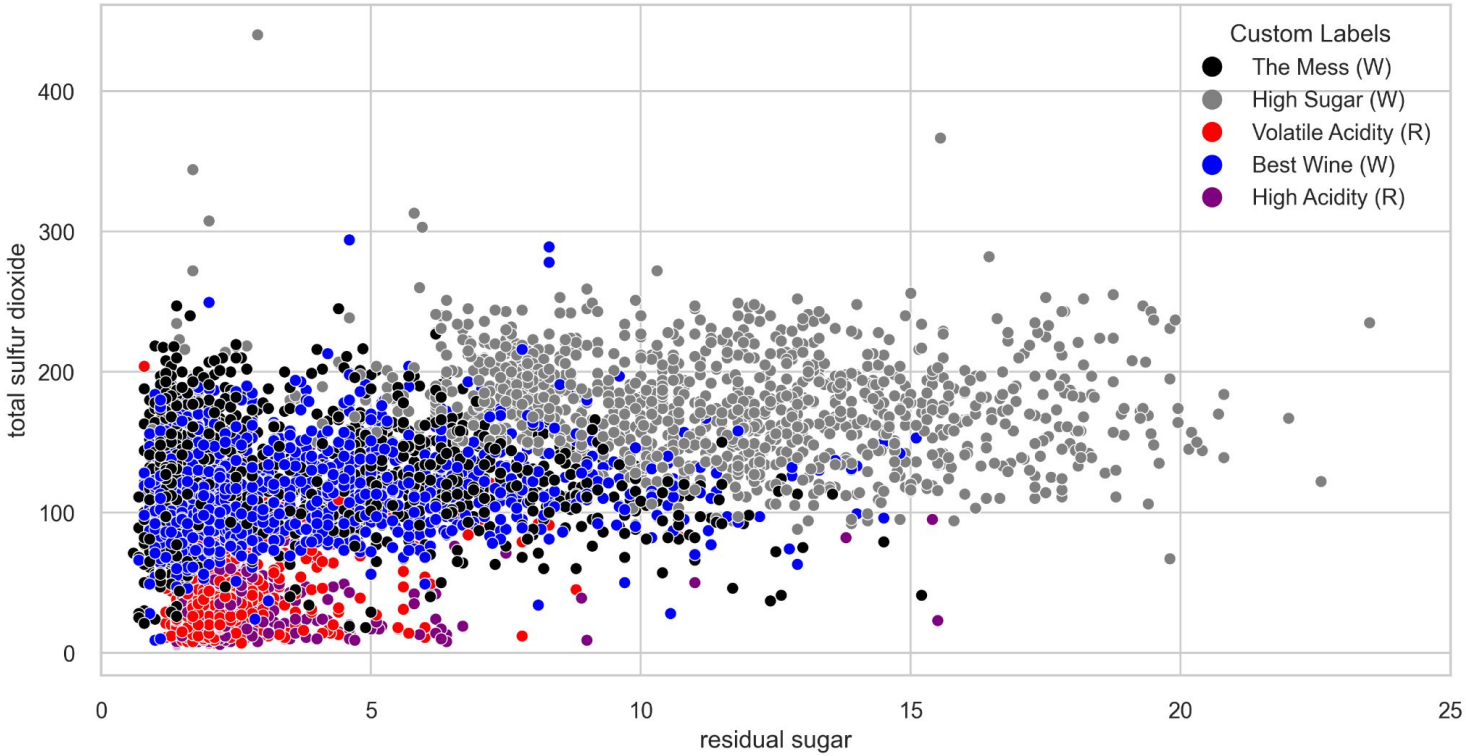




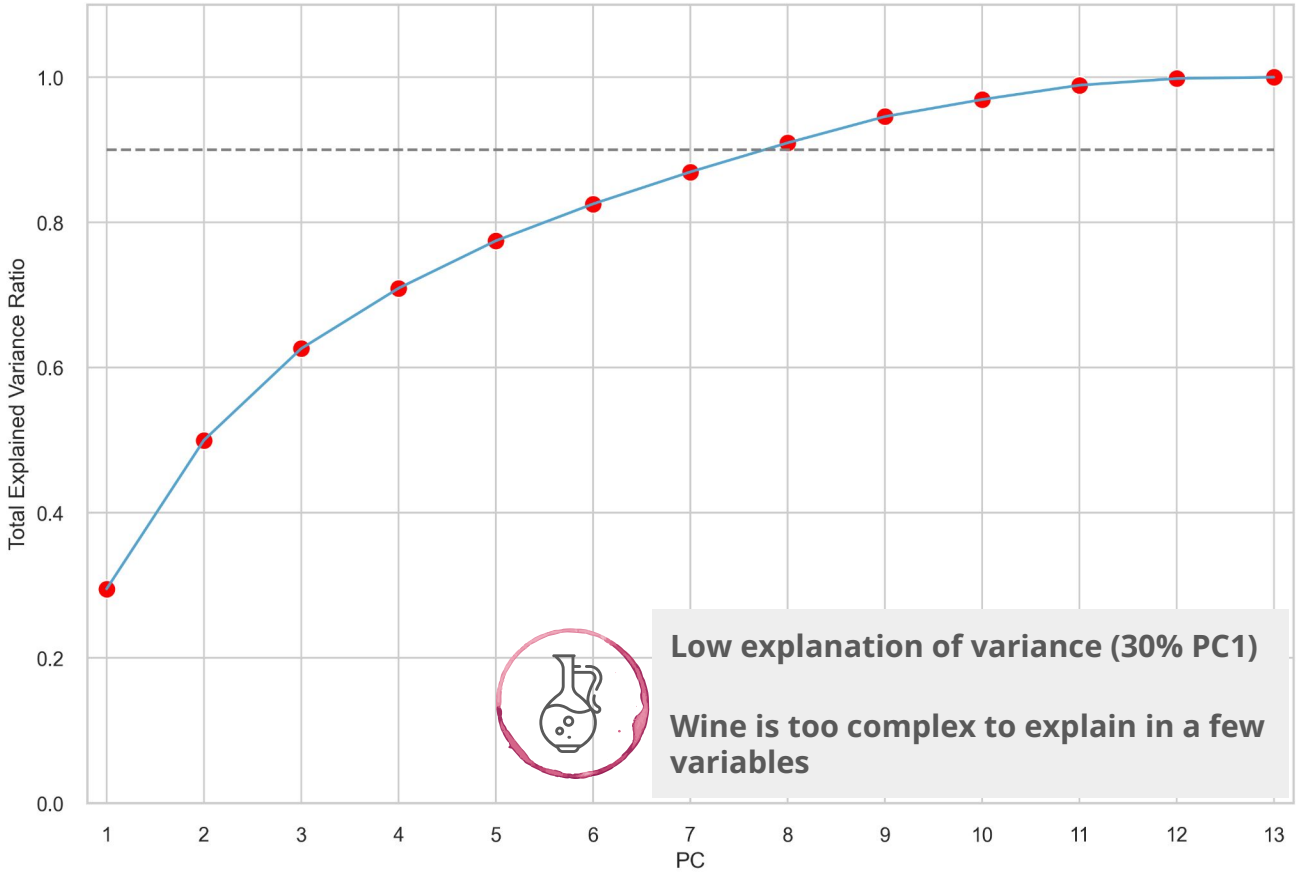
Sugar and Sulfur

Higher preservatives and sugar for white wine

Removing skin -> oxidation -> old wine/opened wine taste



PCA



Low explanation of variance (30% PC1)

Wine is too complex to explain in a few variables

Conclusion

- **Red and white wine clear differences**
  - Clustering and mean scores (r/w)

- Wine is **highly complex** and taste and quality comes from wide range of chemical compounds (PCA weak)

- **Highest alcohol content is better for taste**
  - Bearer of taste and stability

- **The highest quality rated wine has very balanced chemical values and high alcohol content**
  - Balance Index?

Challenges

Bottom up

Look at number of variables in each cluster

KMeans Random

New results each time you run it

PCA

Conceptually hard
What does PC1 mean?