

Group 4

Wine Reviews



Our Team:

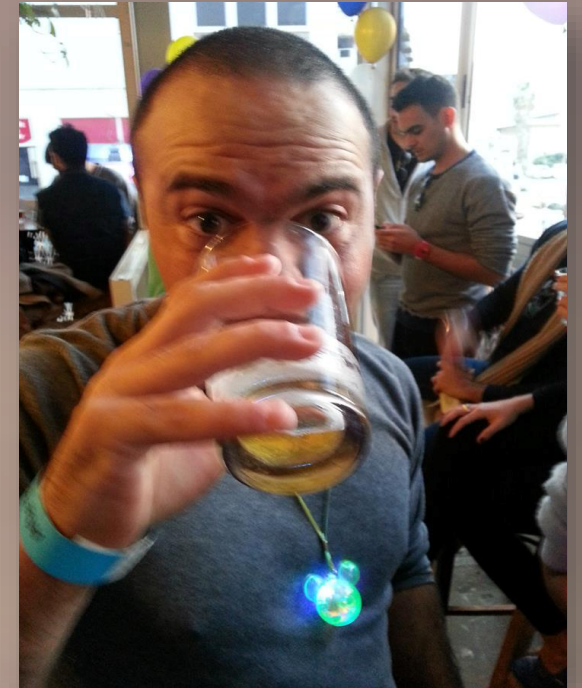
Dor Heldshtein



Lior Rosen



Rotem Zaarour



Taster's Choice...



**Can you describe these wines
taste, flavor, smell?**

Can you estimate their Price?

What would be their Quality?



Agenda:

- ☐ Taster's Choice
- ☐ About the Data
- ☐ Business Questions
- ☐ The Taste of the Results
- ☐ Business Insights
- ☐ Conclusions
- ☐ What's next?



Quick glance at the DATA (130K Wine Reviews):

Variable Name	Country	Description	Designation	Points	Price
Description	Wine origin	Professional review describing the wine's taste, smell, look, etc.	Vineyard designation	Rating (by "Wine Enthusiast")	Cost per bottle
Type	character	character	character	Integer	Integer
Example value	Spain , one of 48 types	description	Acromion, one of 30503 types	80-100	4-2300

Variable Name	Region_1	Region_2	Winery	Variety	
Description	Winery Province/State	Sub-region	Winery name	Grape name	
Type	character	character	character	character	
Example value	Napa Valley	Sonoma	Ponzi	Pinot Noir	

Data source : <https://www.kaggle.com/zynicide/wine-reviews/data>

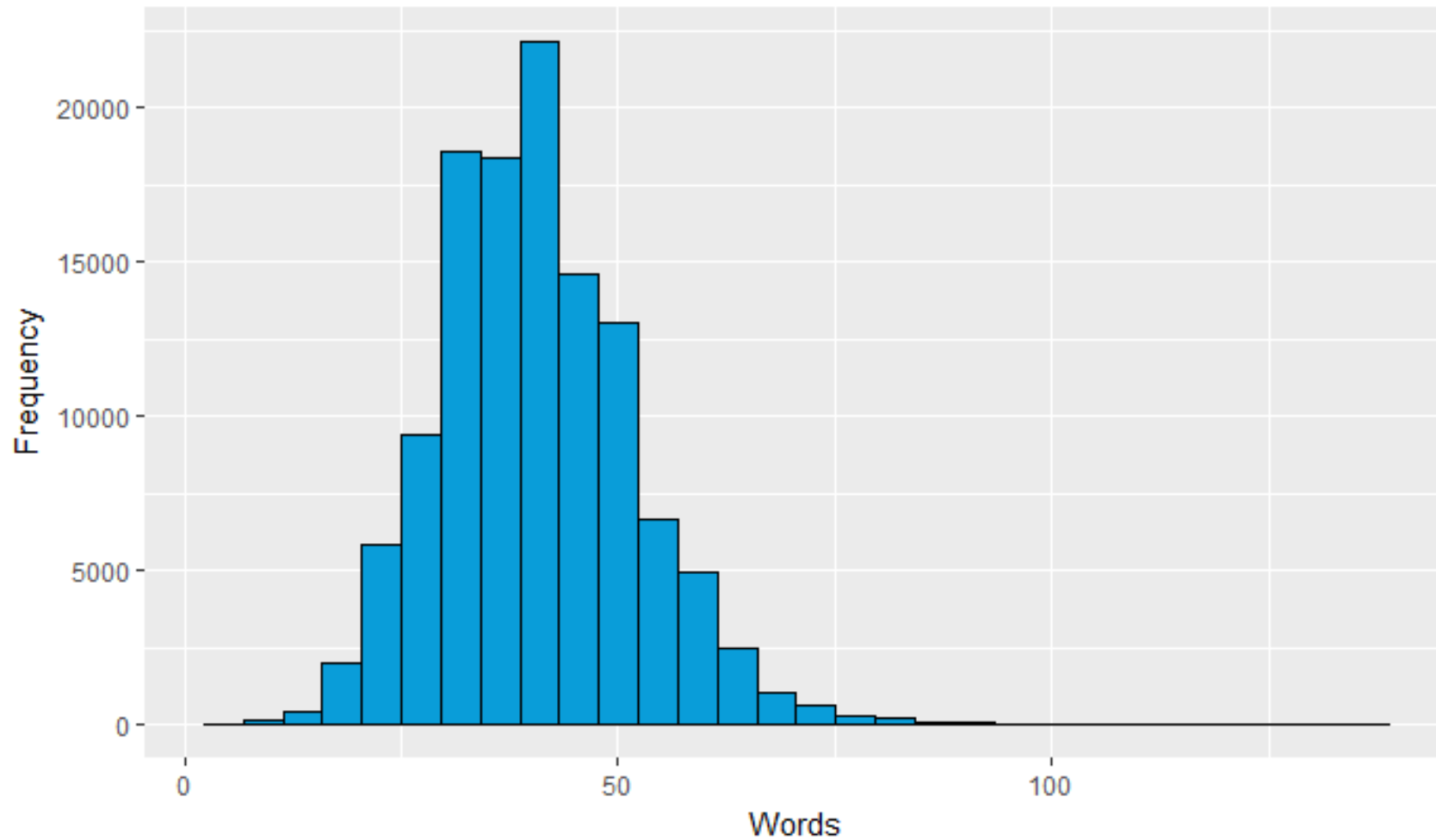
A close-up photograph of a dark glass bottle pouring red wine into a clear, rounded glass. The wine is captured mid-pour, creating a dynamic stream and bubbles. The background is a rustic wooden surface with visible grain and texture. The lighting is warm and focused on the glass and the liquid.

The business questions:

- ☐ Can the data be used in order to estimate the market price of wines?
- ☐ Can this data be used to make wine marketing more efficient and profitable?

EDA:

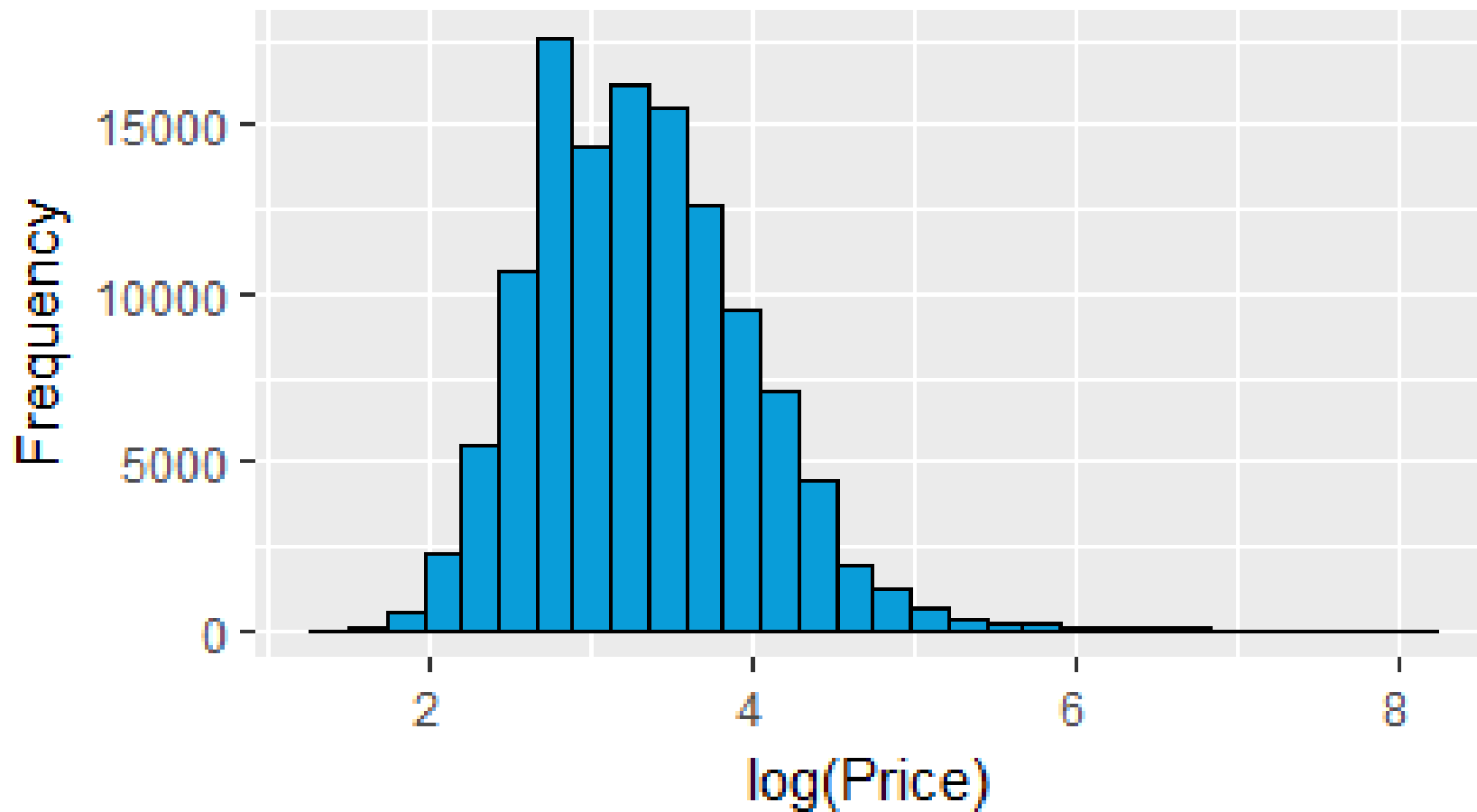
Wine Reviews - Description words amount Histogram



Min.	1st Qu.	Med.	Mean	3rd Qu.	Max.
3	33	40	40	47	135

EDA:

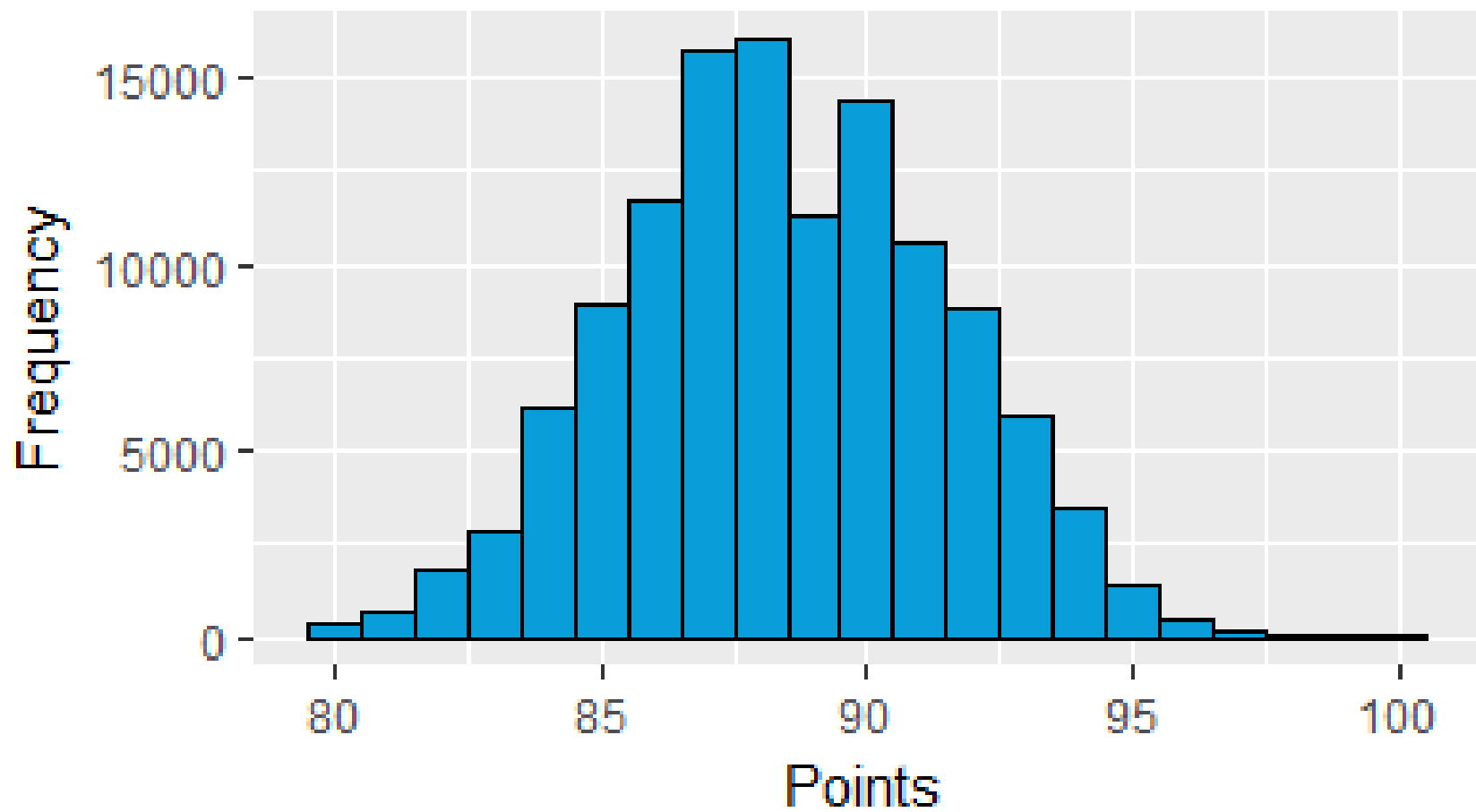
Wine Reviews - Log(prices) Histogram



Min.	1st Qu.	Med.	Mean	3rd Qu.	Max.
1.4	2.8	3.2	3.3	3.7	8.1

EDA:

Wine Review - Points histogram



Min.	1st Qu.	Med.	Mean	3rd Qu.	Max.
80	86	88	88.42	91	100

Top 10 frequent wine varieties in the DATA :

Pinot Noir 12787

Chardonnay 11080

Cabernet Sauvignon 9386

Red Blend 8476

Bordeaux-style Red Blend 5340

Riesling 4972

Sauvignon Blanc 4783

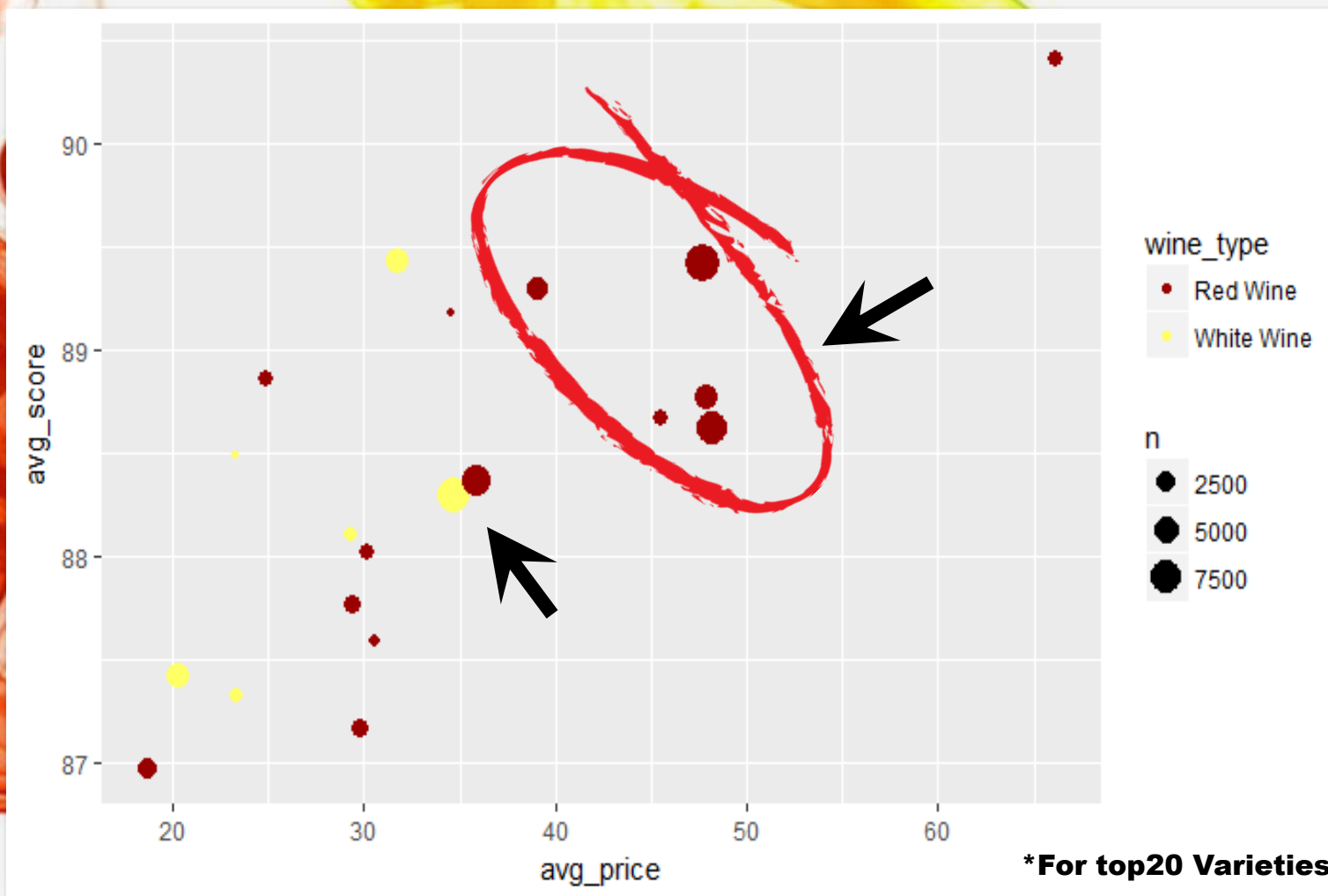
Syrah 4086

RosÃ 3262

Merlot 3062



Red wine Vs. White wine:

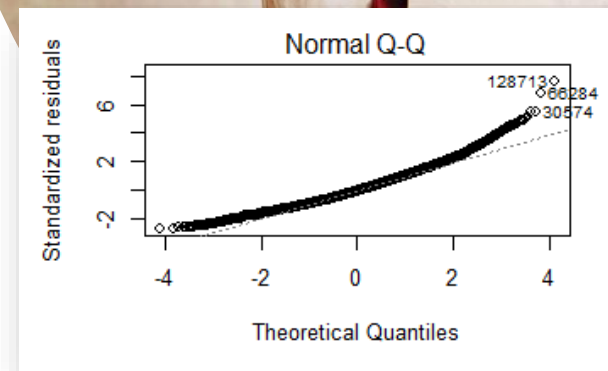
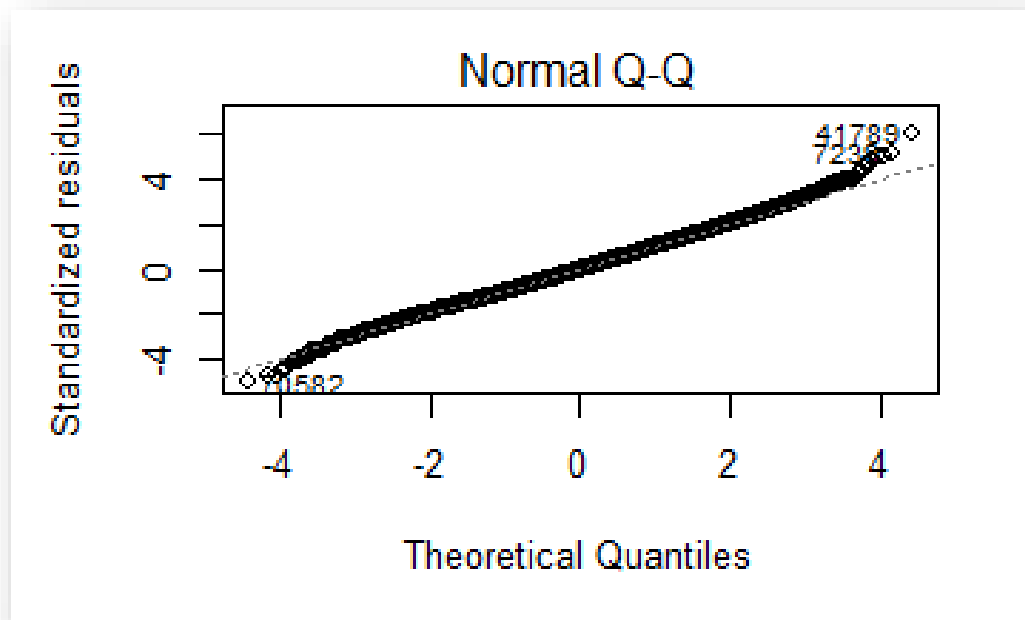


Correlations:

	points	price	desc_nw
points	1.000	0.611	0.514
price	0.611	1.000	0.312
desc_nw	0.514	0.312	1.000

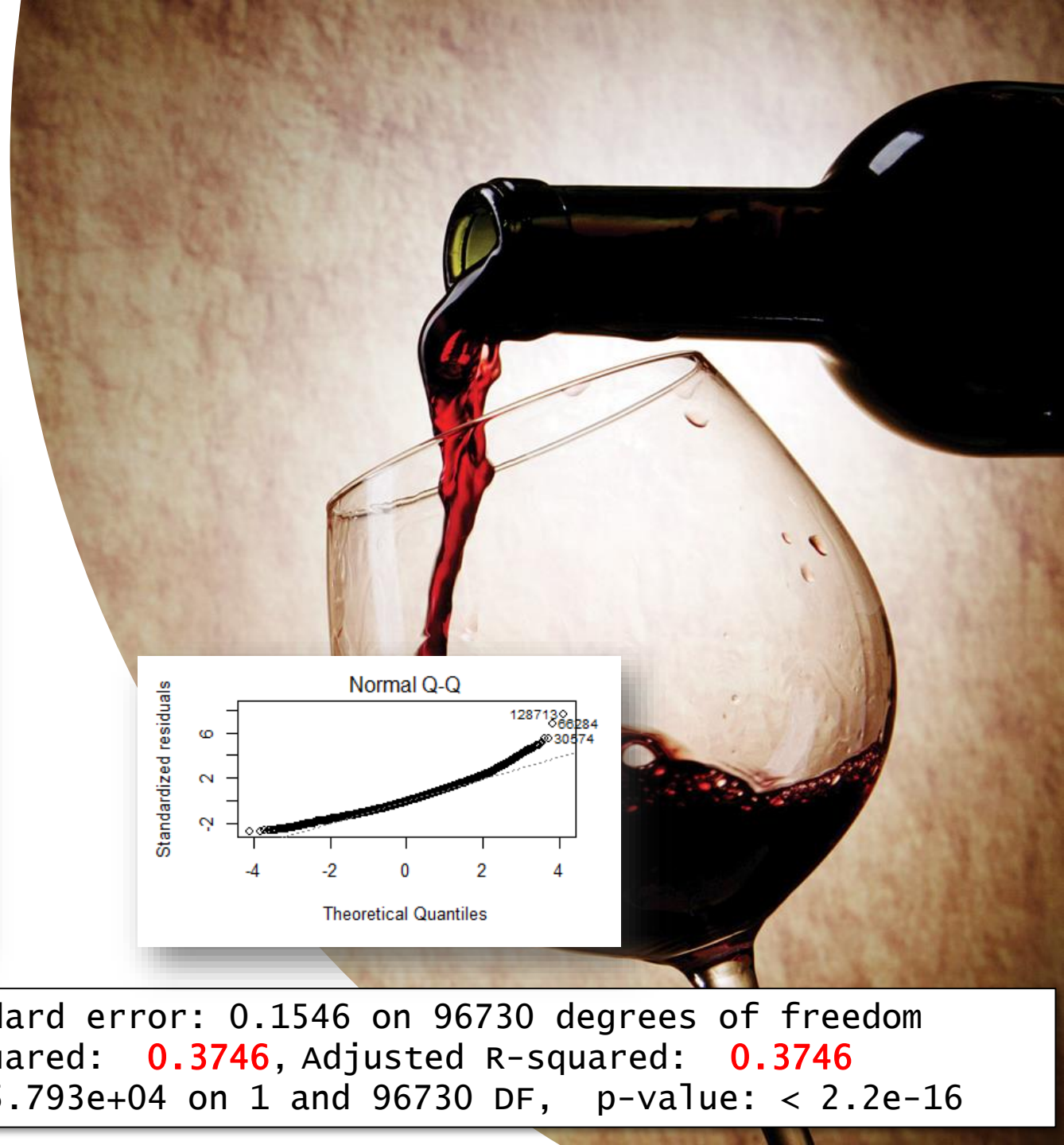
Finding #1

Price can be predicted by wine's rating



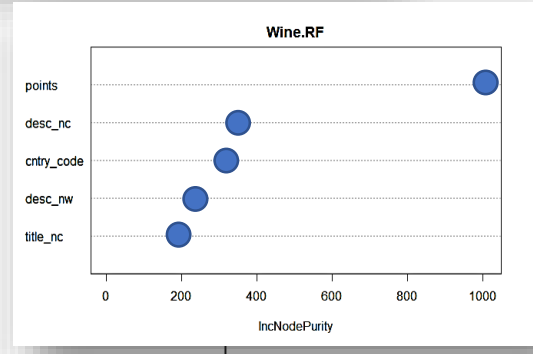
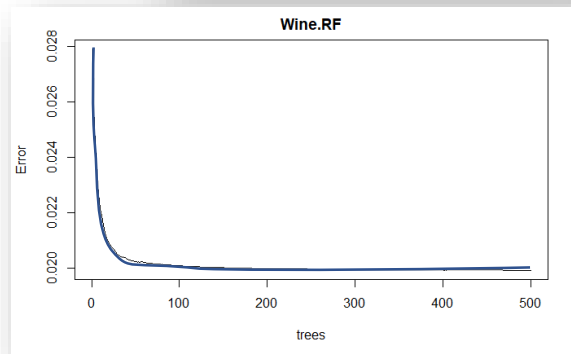
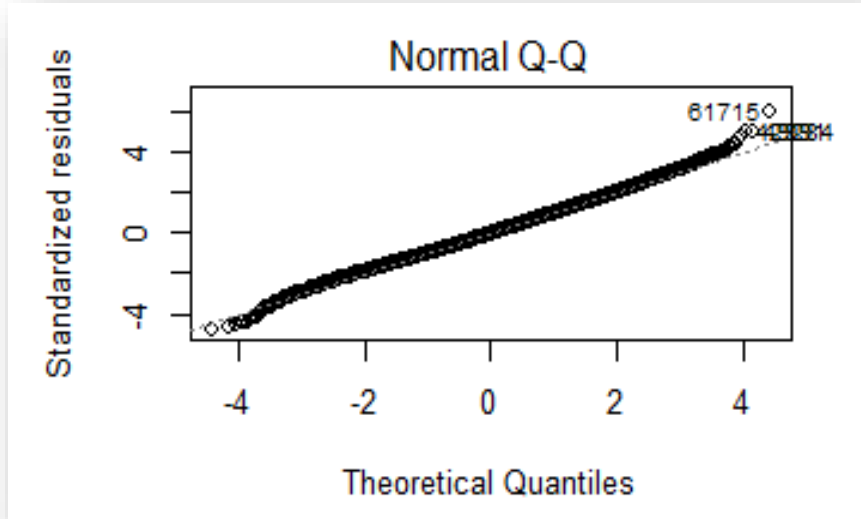
Residual standard error: 0.1546 on 96730 degrees of freedom
Multiple R-squared: **0.3746**, Adjusted R-squared: **0.3746**
F-statistic: 5.793e+04 on 1 and 96730 DF, p-value: < 2.2e-16

`lm(log(price) ~ points, data = clean.train)`



Finding #2

Word count in a review is also a predictor of wine's rating ratio



```
(formula = log(price) ~ ., data = clean_train)
```

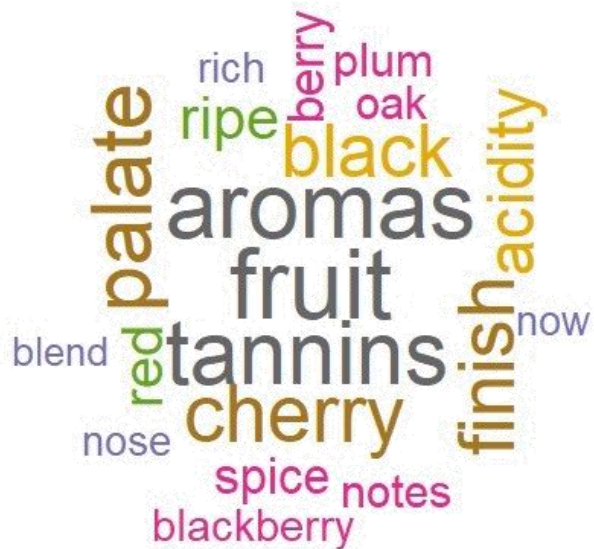
red residuals: 0.01992124
Var explained: 47.95



Finding #3

Specific words appearing in the wine review can predict wine's variety

Red Wines



White Wines



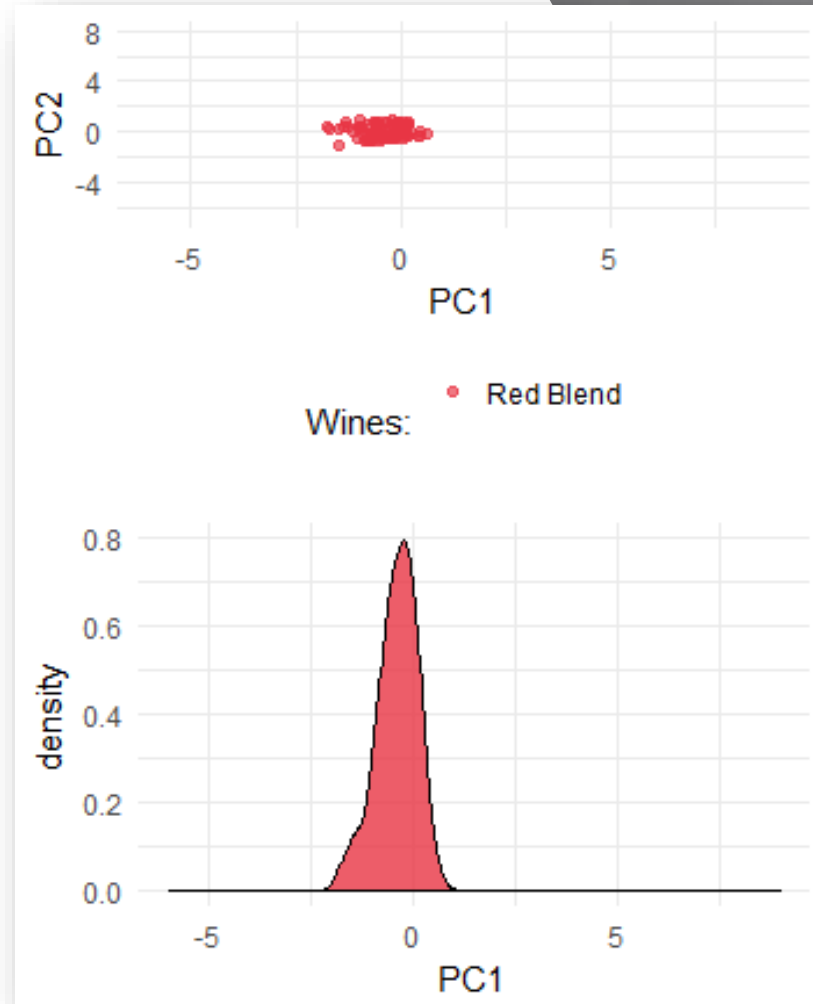
Clustering “Red Blend” Wines

word <fctr>	freq <dbl>
aromas	888
blend	795
palate	756
black	739
tannins	736
cherry	700
cabernet	562
fruit	518
red	484
finish	473
berry	429
spice	416
plum	404
sauvignon	389
merlot	373
ripe	337
syrah	331
blackberry	324
pepper	313
notes	292



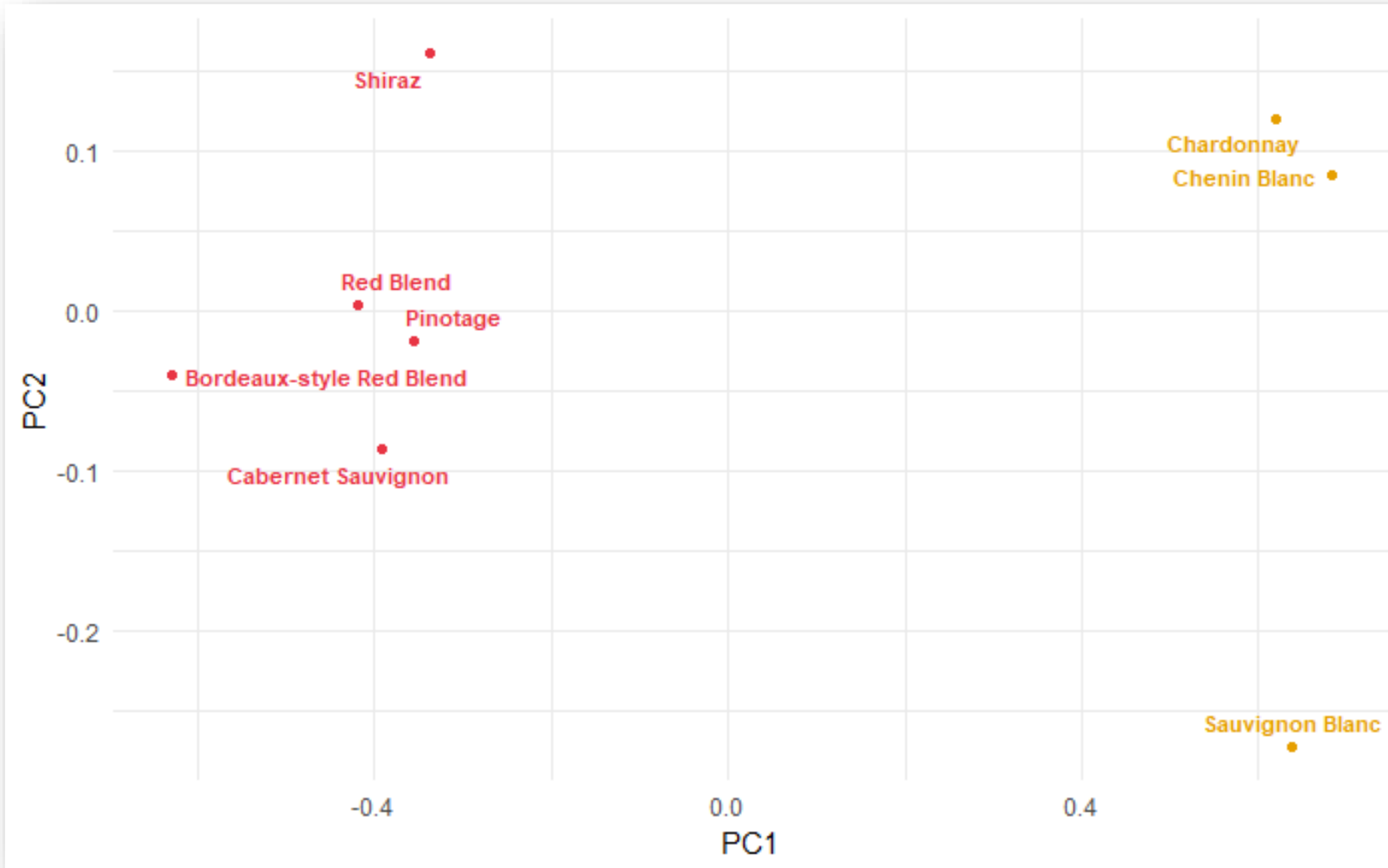
Segment – red blend

Principal Component Analysis (PCA) for Red Blend Wines



Segment – red blend

PCA for Different Wine Varieties



Segment – South Africa

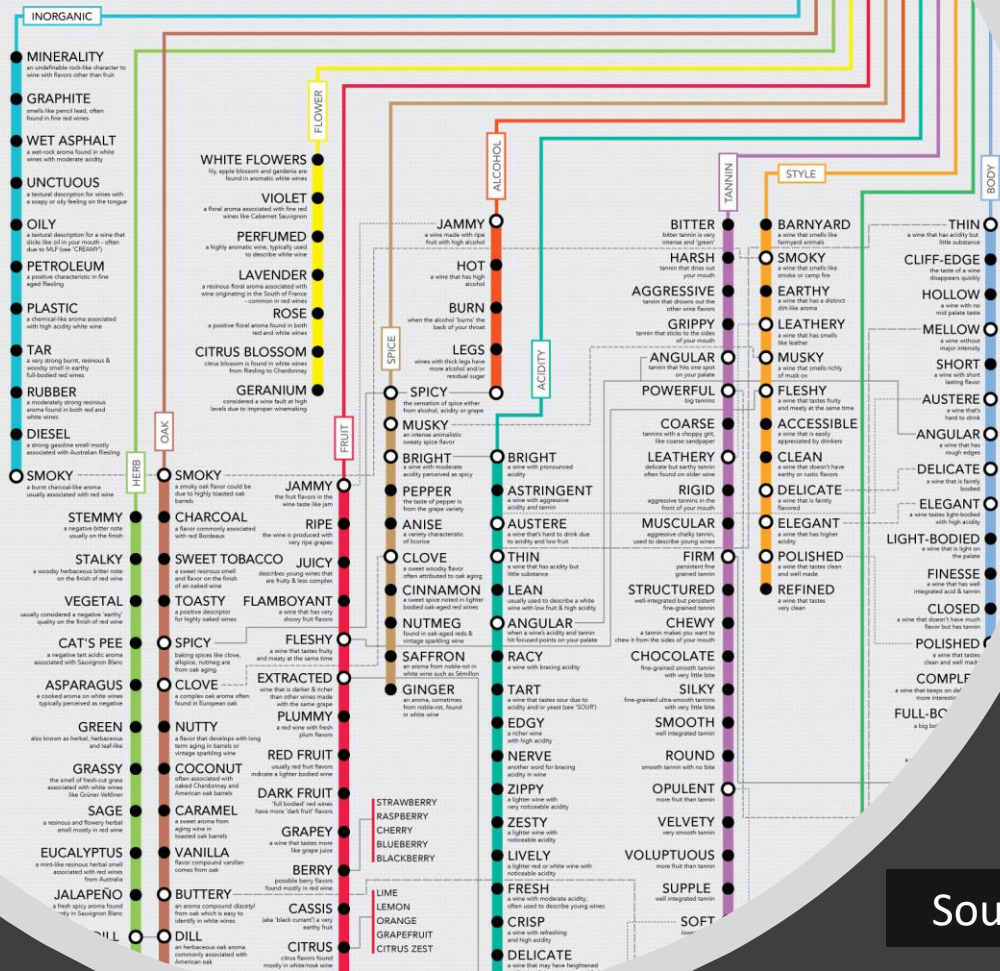


WINE DESCRIPTIONS

& WHAT THEY MEAN

wine descriptions can be divided into
TWELVE CATEGORIES

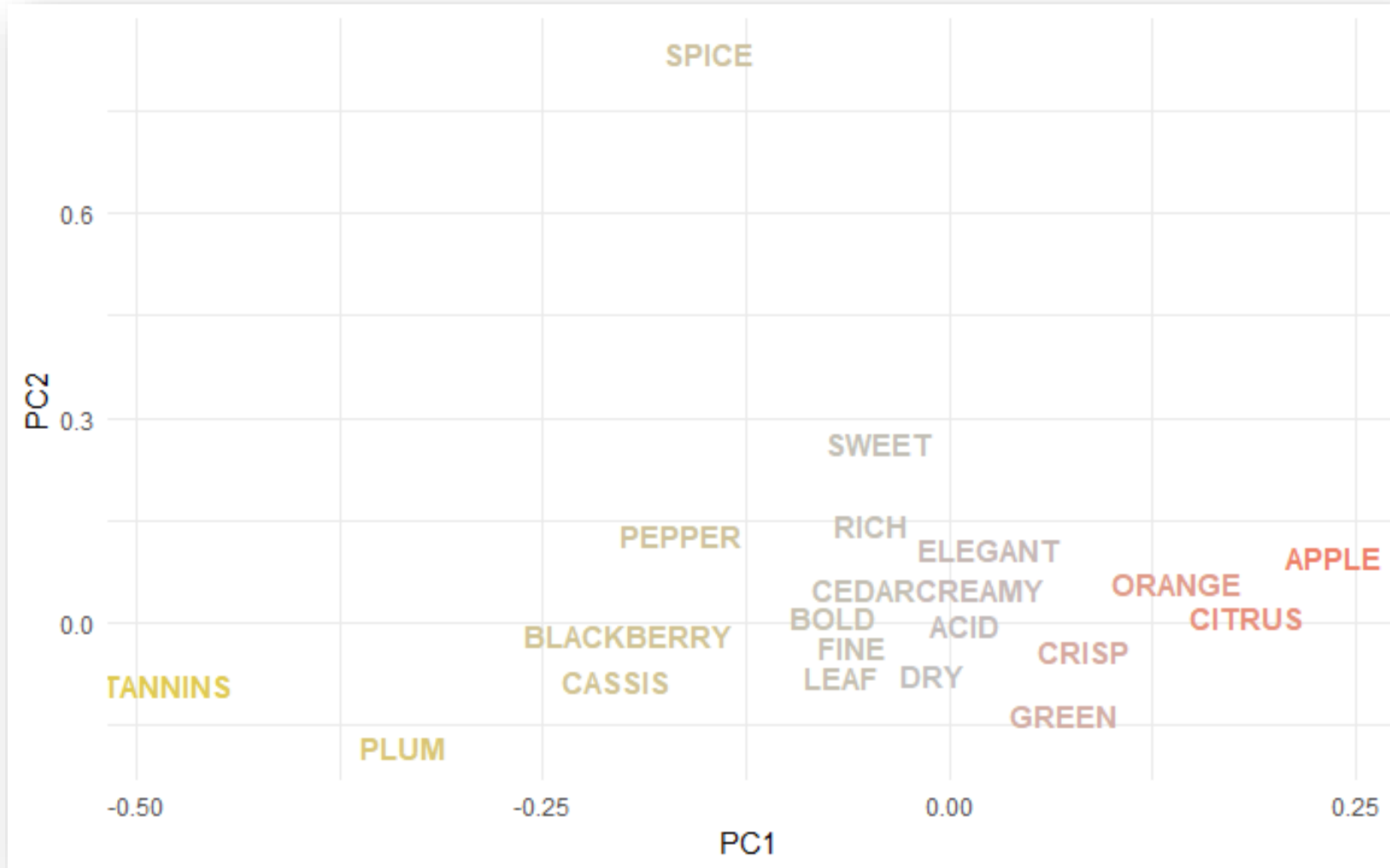
BODY
YEAST
STYLE
TANNIN
ACIDITY
ALCOHOL
SPICE
FRUIT
FLOWER
HERB
OAK
INORGANIC



Some of the professional description words of wine...

Source: Wine FOLLY

Stem words



Segment – South Africa



Business Insights and Applications

The presented data enables to **predict a wine's value-for-money** based on **characteristics of its review** (length of review, presence of specific words).

The longer the review is, the more stem words are used to describe the wine, the higher its quality is.

Various clusters have been identified, based on words used for the wine's description, wine variety and its geographical origin.

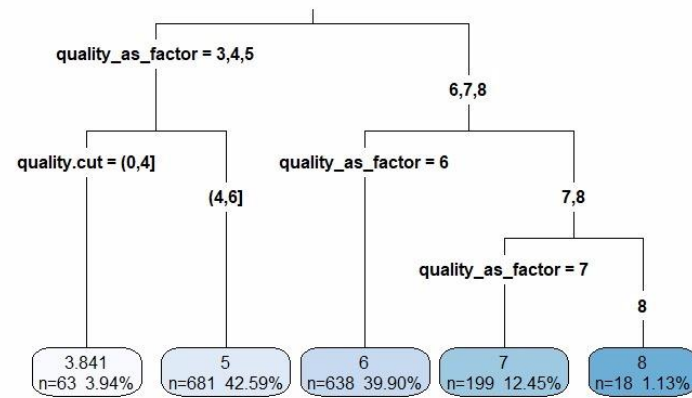
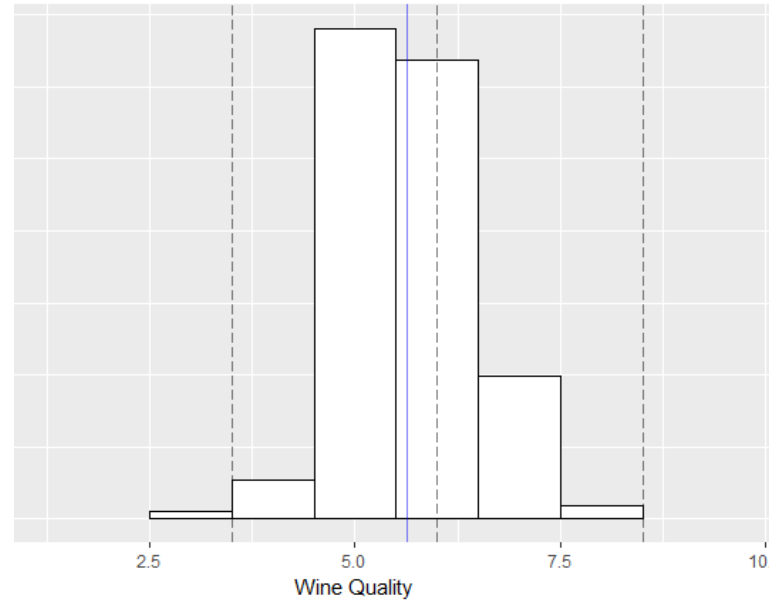
The data can be used to define useful marketing buzzwords in order to describe wines in a way that will highlight its taste and quality.

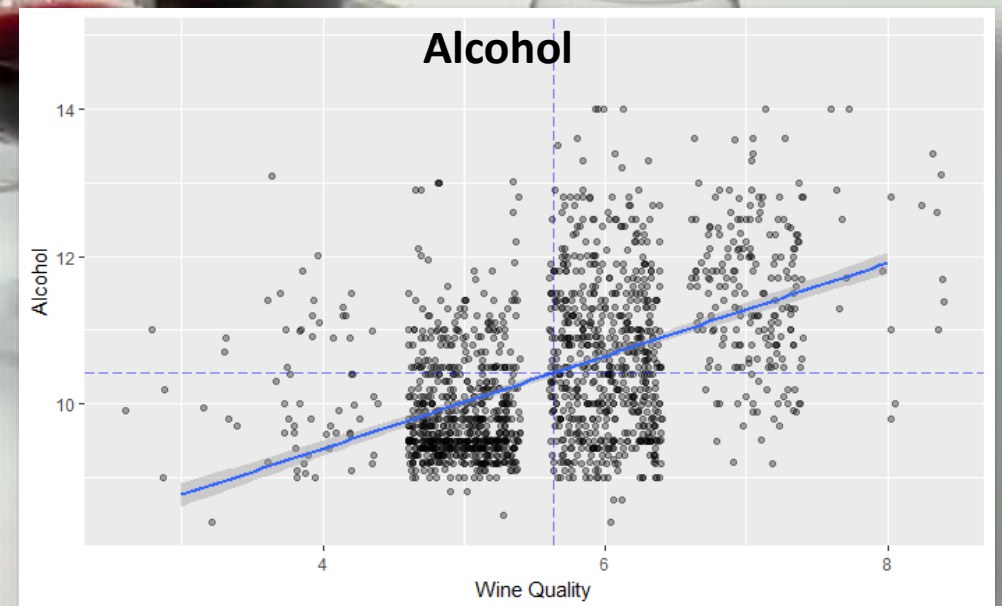
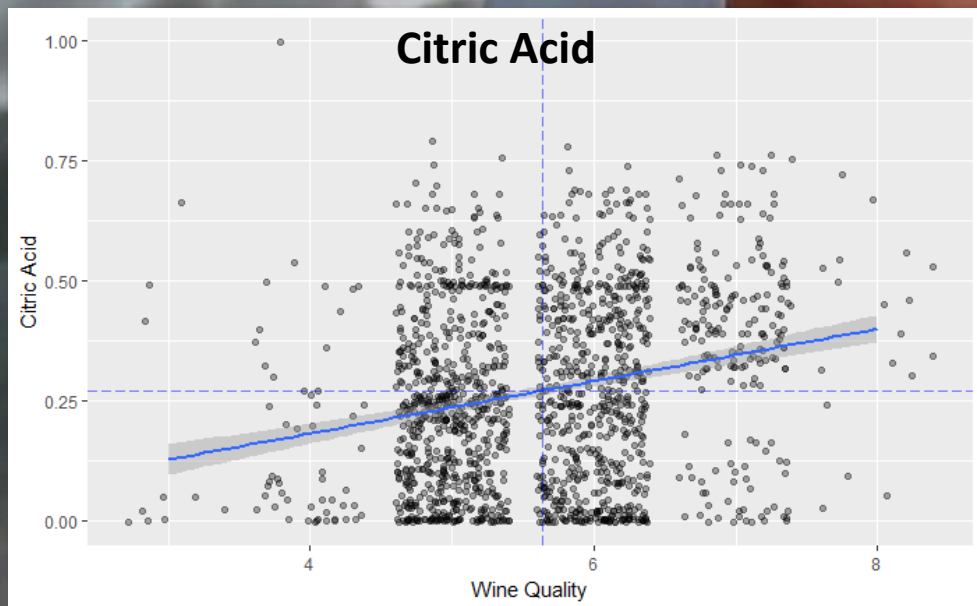
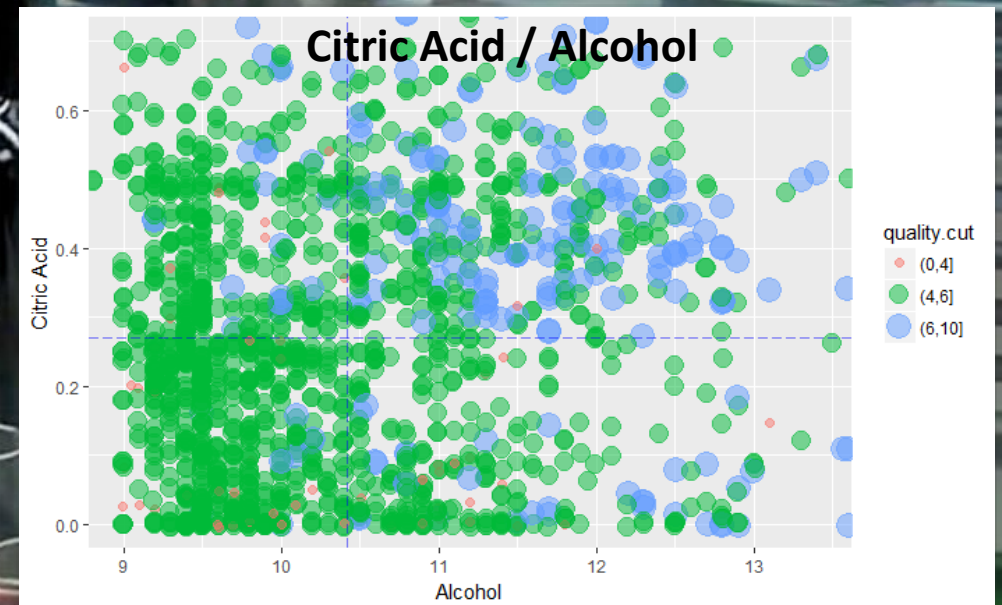
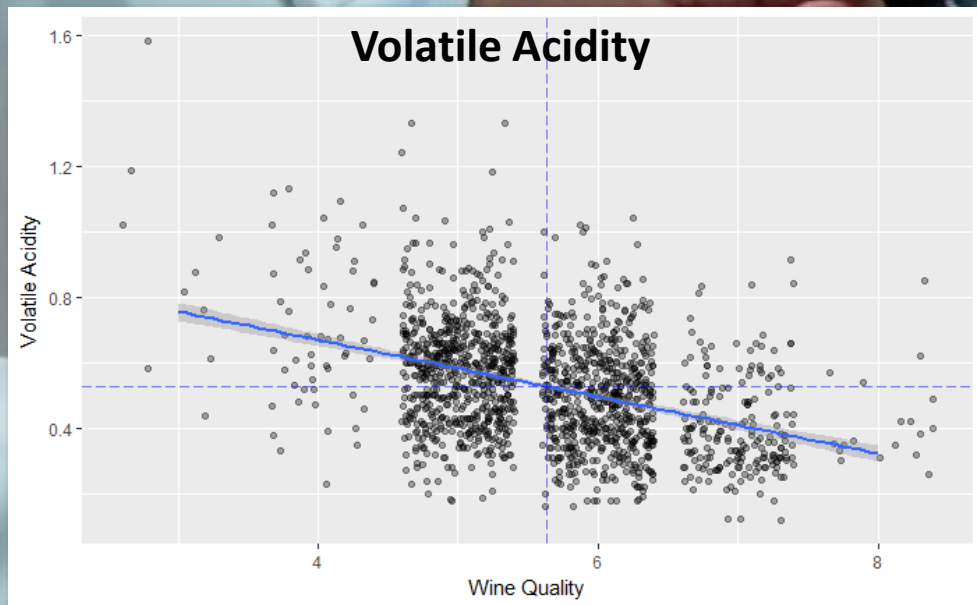
Main application - advertising



Enriching the data

- In order to enrich the data with additional qualitative parameters of wines, we have added a data of 6500 white and red Portuguese Wines
- Quality ranked 0-10, as a factor of :
 - Alcohol %
 - Citric Acid %
 - Volatile Acidity
 - pH
 - Chlorides
 - Density
 - Sulphates
 - Total sulfur dioxide
 - Residual sugar







The end