

Wine Reviews Analysis

Matt Thomas, Kevin Finitiy,
Kevin Lennon, Jae Hyun Lee

Dataset consists of 130,000 wine reviews from Wine Enthusiast

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

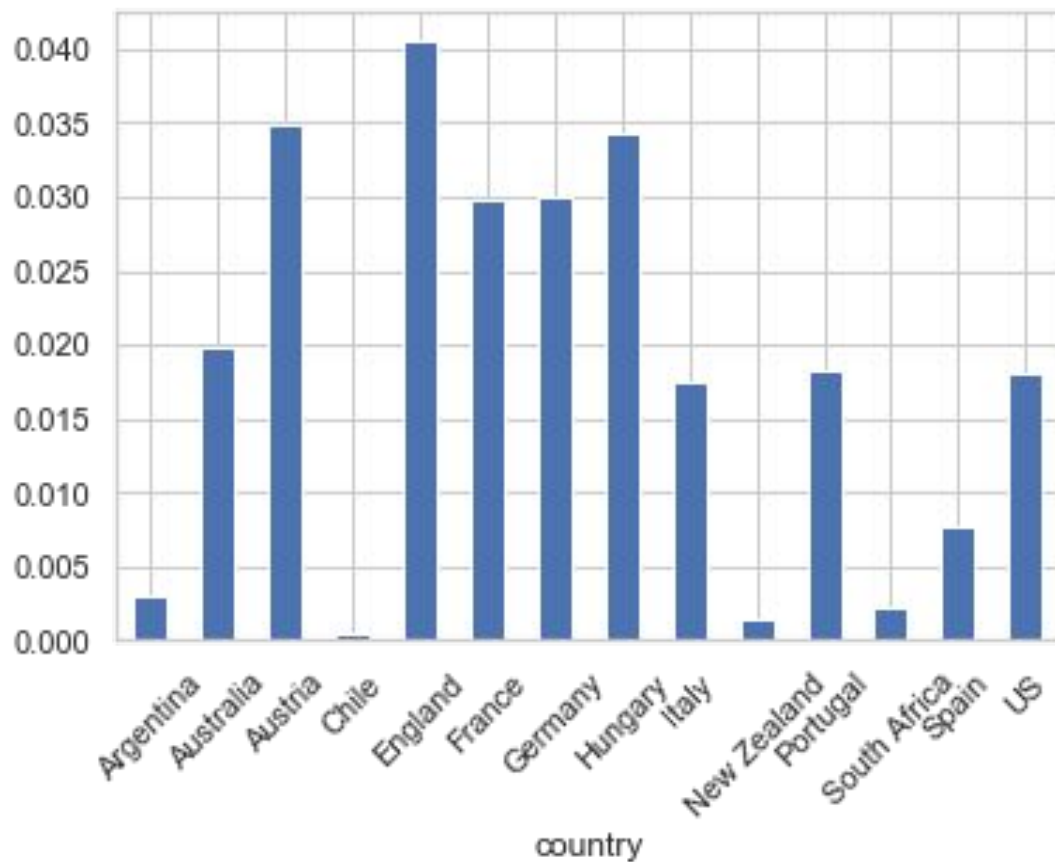
Columns Include:

- Country: country wine was made
- Description: tasting notes
- Designation: vineyard name, if applicable
- Points: points awarded by taster, out of 100
- Price
- Province: general winemaking area, e.g. Virginia
- Region: two columns to allow for subregions, e.g. Russian River Valley, Sonoma
- Title: Wine name, if applicable
- Variety
- Winery

Goals

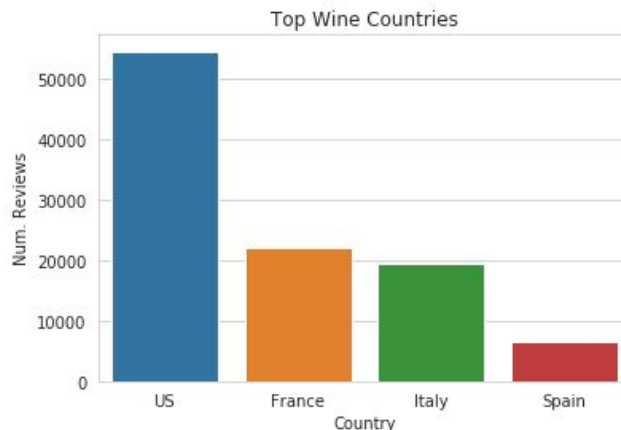
- Using points as a measure of quality, what can we learn about a wine's features that influence its rating?
 - Where a wine is made?
 - Who tastes the wine?
 - Price?
- General Data Analysis: what insights can we glean from the dataset?
- Text Analysis: how to descriptors used in the wine's tasting notes affect its rating/quality?

Countries with the most “Superlative” Ratings



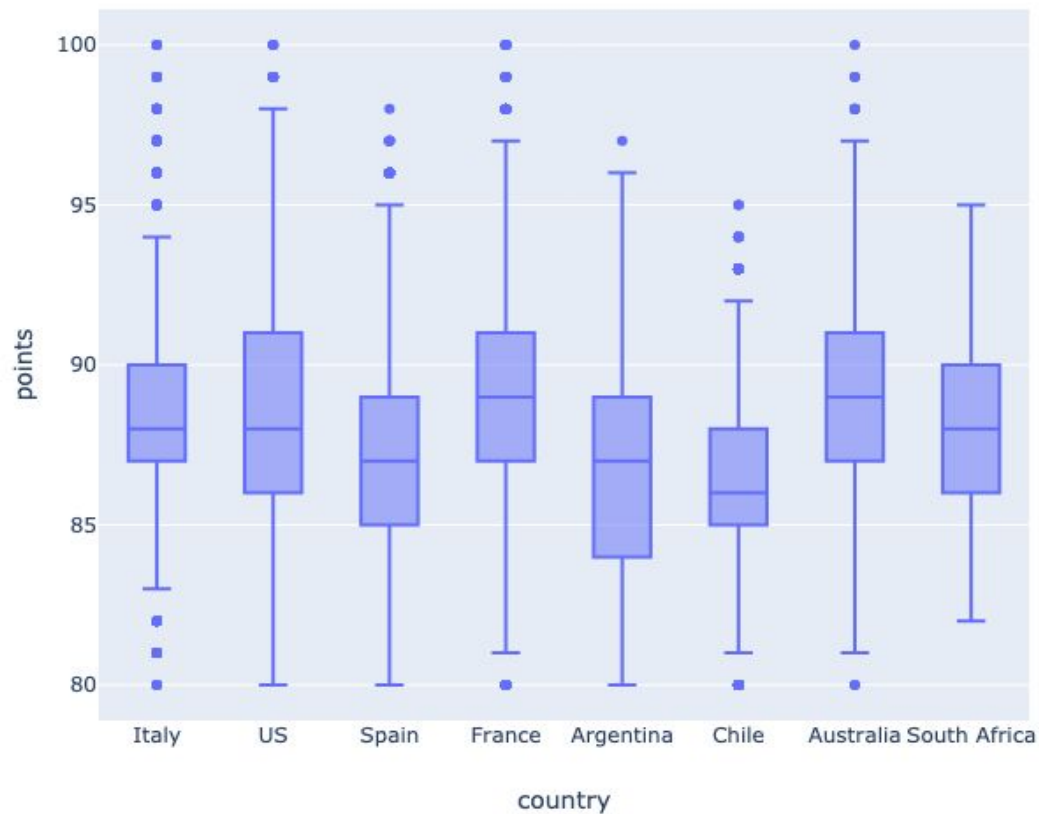
A comparison of top wine producing countries:

- United States
- France
- Italy
- Spain
- Australia
- Argentina
- Chile
- South Africa

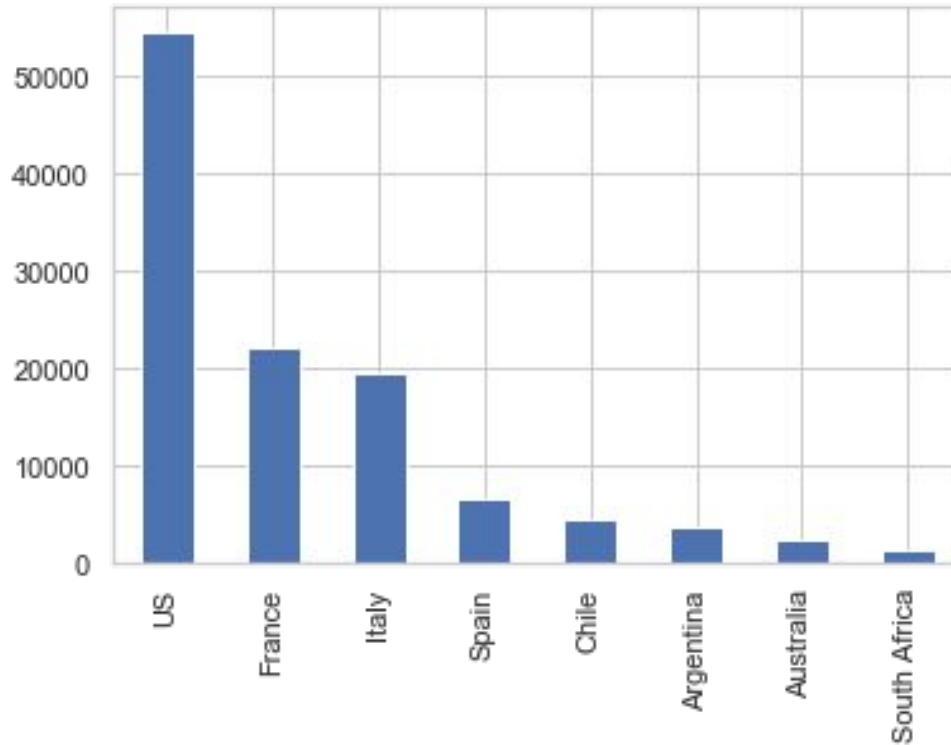


Why these countries: Because these are the top wine producers that also make a wide variety of wine types and styles: they make up a majority of what you would find in a wine store.

Points Statistics for Top Wine Producing Countries



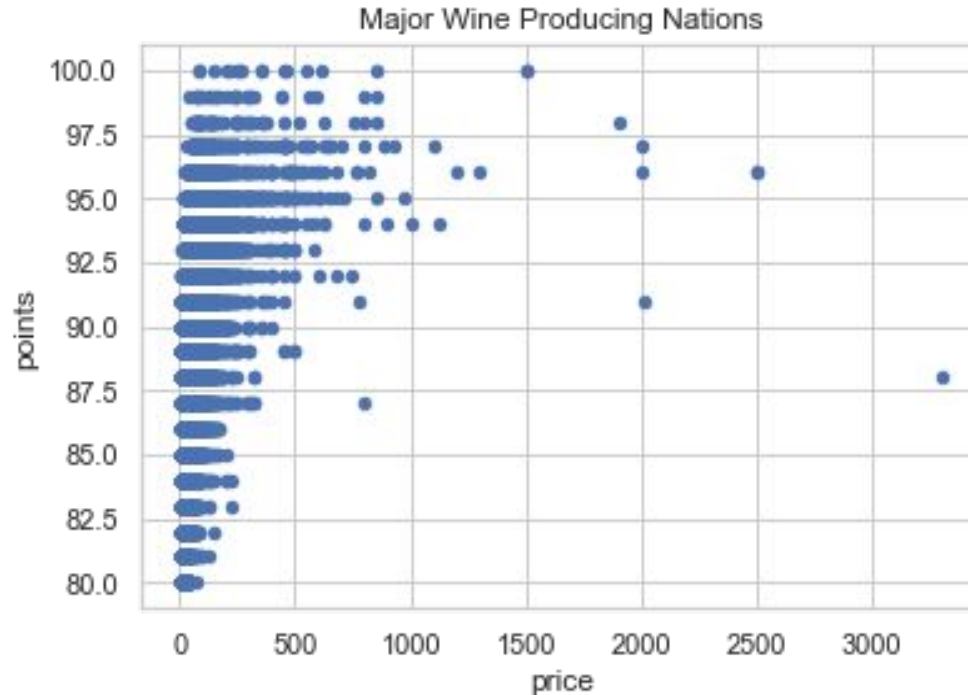
Need to keep in mind number of review per country



We can query the data to see which wines are rated highest by country:

	US :	winery	points	price
	45798	Cardinale	100	200.0
	France :	winery	points	price
	36528	Krug	100	259.0
	Italy :	winery	points	price
	7335	Avignonesi	100	210.0
	Spain :	winery	points	price
	40307	Emilio Moro	98	450.0
	Australia :	winery	points	price
	345	Chambers Rosewood Vineyards	100	350.0
	South Africa :	winery	points	price
	9912	De Toren	95	330.0
	Argentina :	winery	points	price
	82754	Bodega Catena Zapata	97	120.0
	Chile :	winery	points	price
	73993	Lapostolle	95	90.0

Generally, the higher the price the more likely the rating is higher, but what are the outliers?



	country	points	price	province	
1558	France	98	1900.0	Bordeaux	Château Margaux 2009 Margaux
15840	France	96	2500.0	Bordeaux	Château Pétrus 2014 Pomerol
65352	France	97	2000.0	Bordeaux	Château Pétrus 2011 Pomerol
80290	France	88	3300.0	Bordeaux	Château les Ormes Sorbet 2013 Médoc
98380	France	96	2500.0	Burgundy	Domaine du Comte Liger-Belair 2010 La Romanée
113564	France	96	2000.0	Burgundy	Domaine du Comte Liger-Belair 2005 La Romanée
120391	US	91	2013.0	California	Blair 2013 Roger Rose Vineyard Chardonnay (Arr...

A Closer Look at Wineries Near UVa

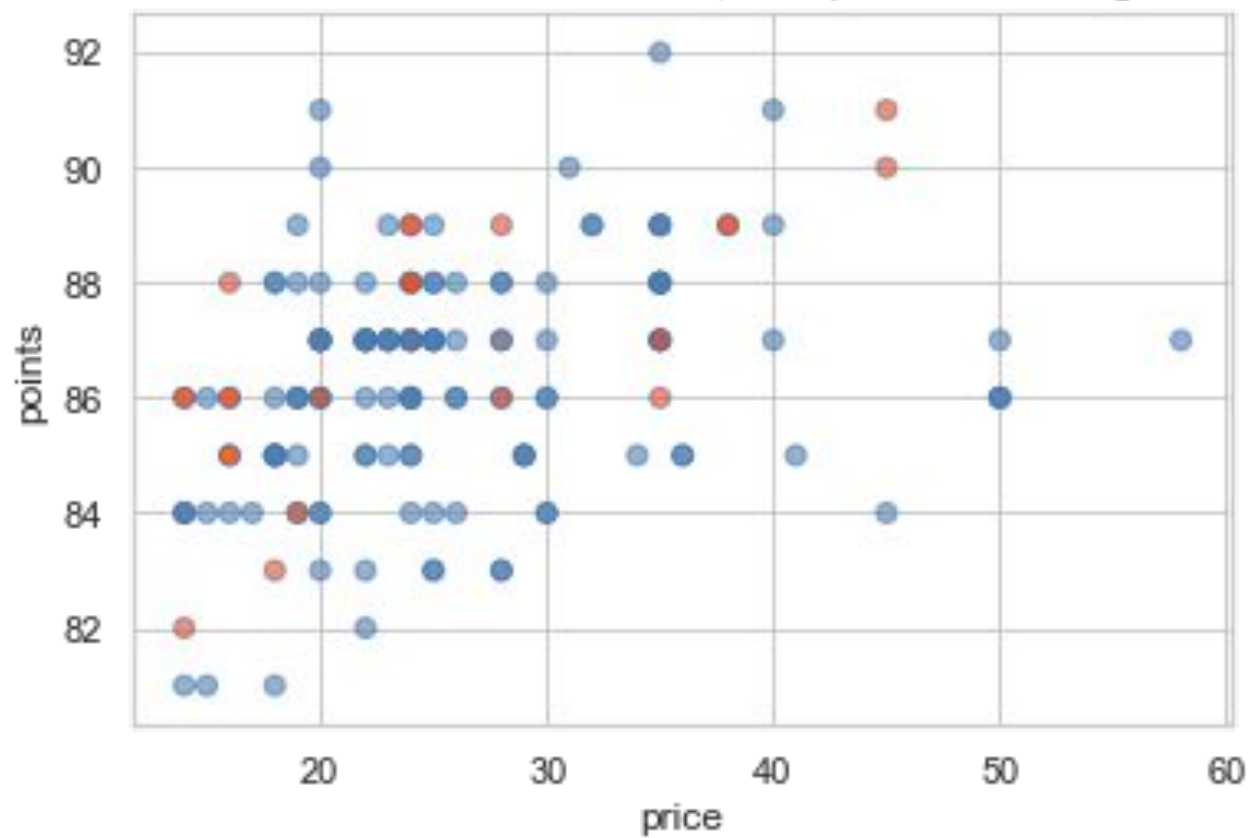
What if you're a data science student (or computer science professor) who wants to visit wineries near you?

It turns out the Charlottesville area has its own regional designation called 'Monticello' and we can query the database to look for it.

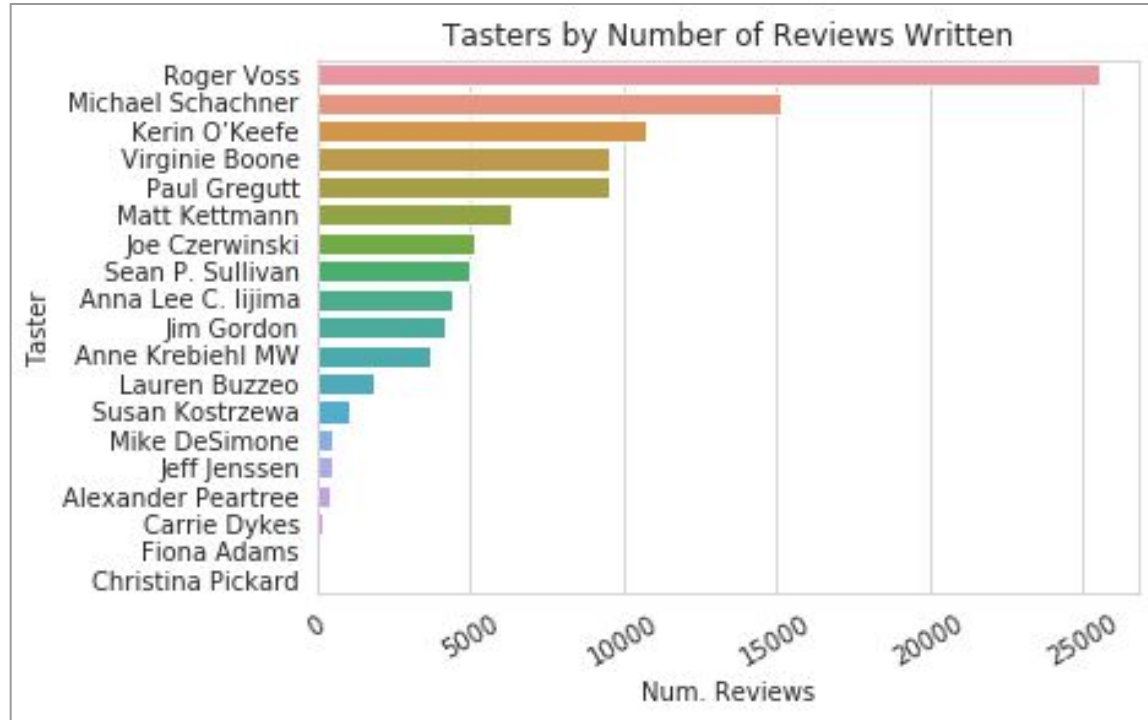
There are three wineries with wines rated above 90 points near UVa: Lovington, King Family Vineyards, and Trump (yes, that Trump).

See Jupyter Notebook for full data query

Charlottesville-Area Wineries, Trump Wines in Orange



The Wine Enthusiast tasters



The Wine Enthusiast tasters

We calculated some aggregate statistics about each taster in the dataset:

- Most-reviewed country
- Highest-rated country
- Most-reviewed varieties
- Highest-rated varieties
- Average price of wine reviewed

Taster Profile

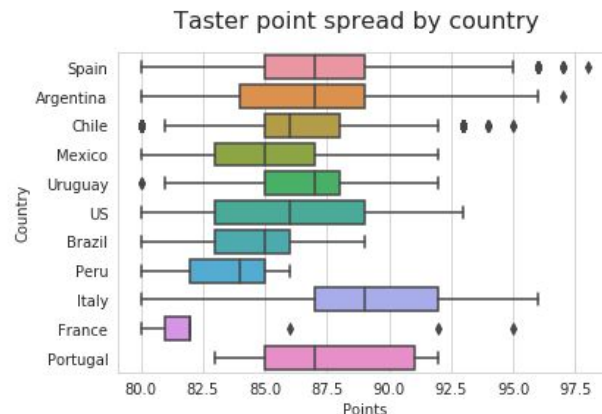
Based on the aggregate taster statistics we collected, we created an interactive function to display a profile for each taster, to indicate their preferences and rating trends.

Taster Profile

Name: Michael Schachner
15134 reviews

Most-reviewed country: Spain

Favorite country: Italy



Top varieties for this taster:

Variety	# Reviews	Average Score	Score Std. Dev.
Malbec	1652	87.417070	3.141296
Red Blend	1496	88.397059	3.007254
Tempranillo	1439	87.492008	3.151005
Cabernet Sauvignon	1358	86.462445	2.728380
Chardonnay	877	84.990878	2.330052

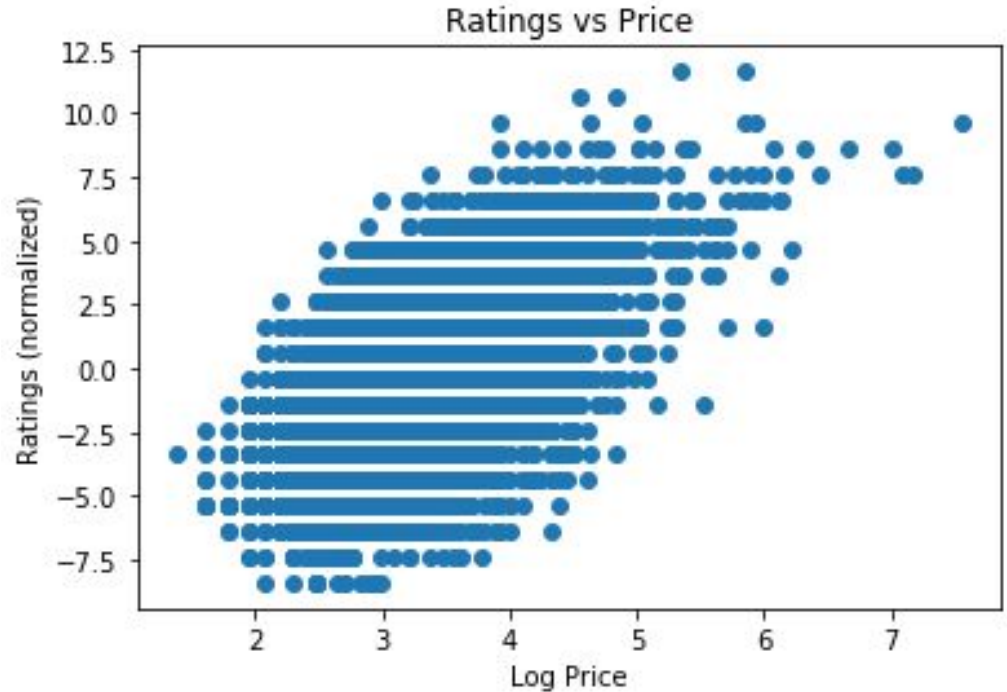
Wine Value Tool

We created a tool to help you find good value wine for the money!

- Most people don't know enough about wine to pick out a good one on their own
- Current wine apps will show you value for a specific bottle you scan
 - Who has time to scan every bottle in the store??
- This tool shows you which traits of wine predict value (country, province, type of wine, descriptions, etc..)

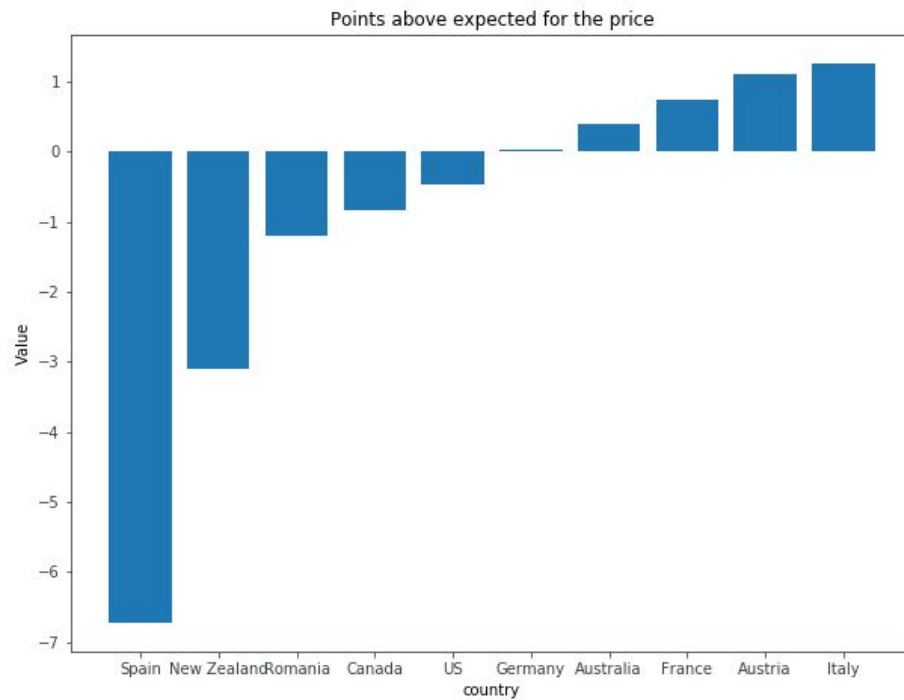
What is a good “Value” Wine?

- Used regression to predict the rating of a wine based on log price
- The residual errors were treated as “Value”
 - Wines above their predicted rating for the price are good value!
- We next looked how to predict value



Factors in predicting Value

- Country, Wine Type, Province, Region, Winery, and words used in the description
- The program lets you drill down to use multiple predictors at once
- The chart to the right shows the best countries for good value Rieslings



Natural Language Processing

- To find which words predict good value, we used the Natural Language Toolkit
- Removed typical stopwords, as well as non-predictive common words in this dataset
- Can be combine with other filters as well

Please select one:

description

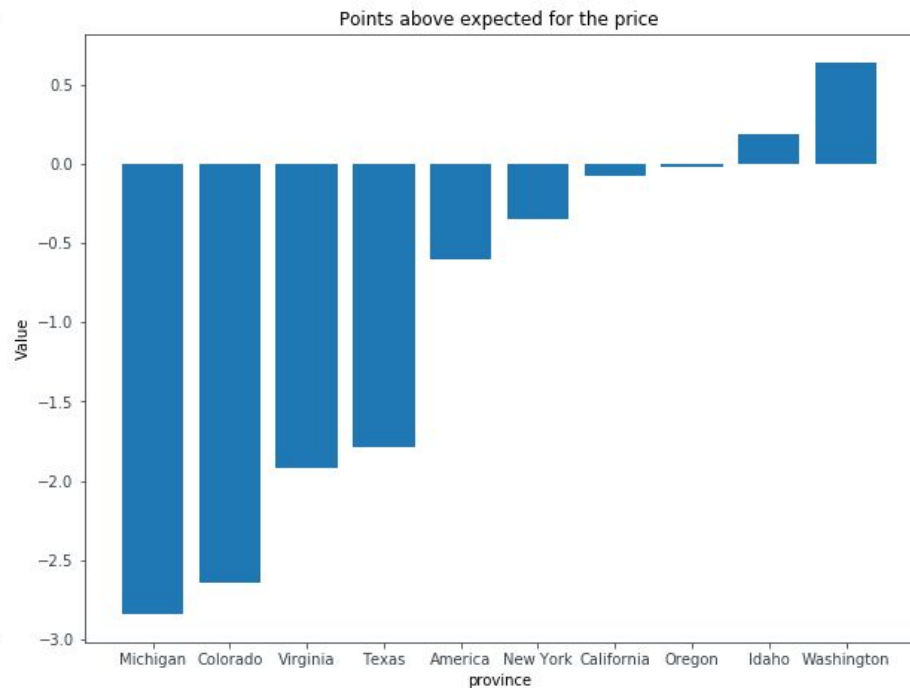
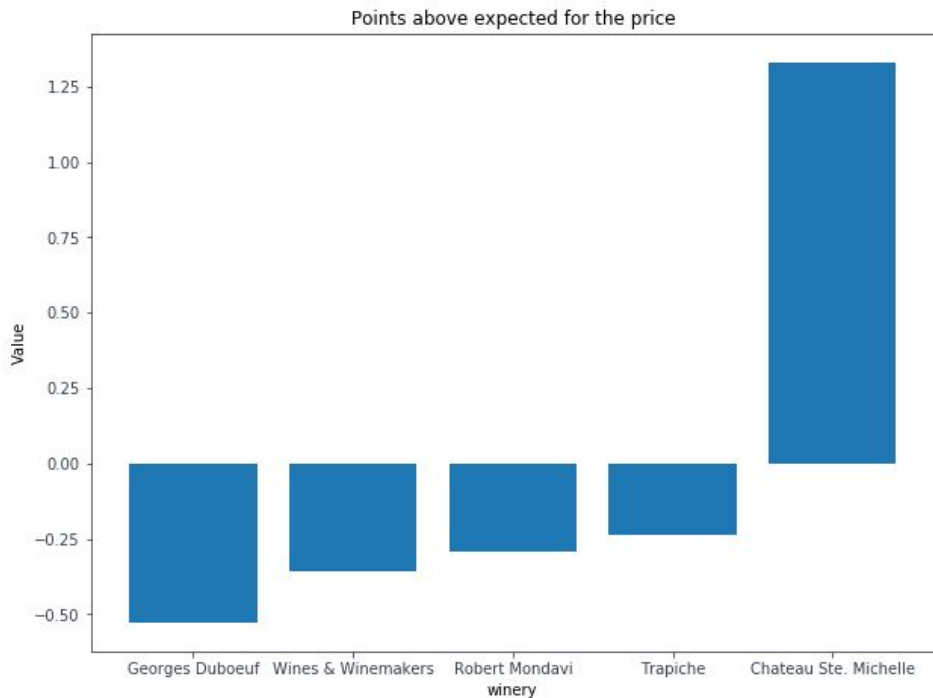
The biggest factors in good value and their average points above the price are:

	value
complex	0.800884
long	0.789865
concentrated	0.695732
cassis	0.624261
full	0.600374

The biggest factors in below average value and their average points below the price are:

	value
along	-0.897819
bit	-0.779204
soft	-0.583661
sweet	-0.474644
oak	-0.436571

Other Examples

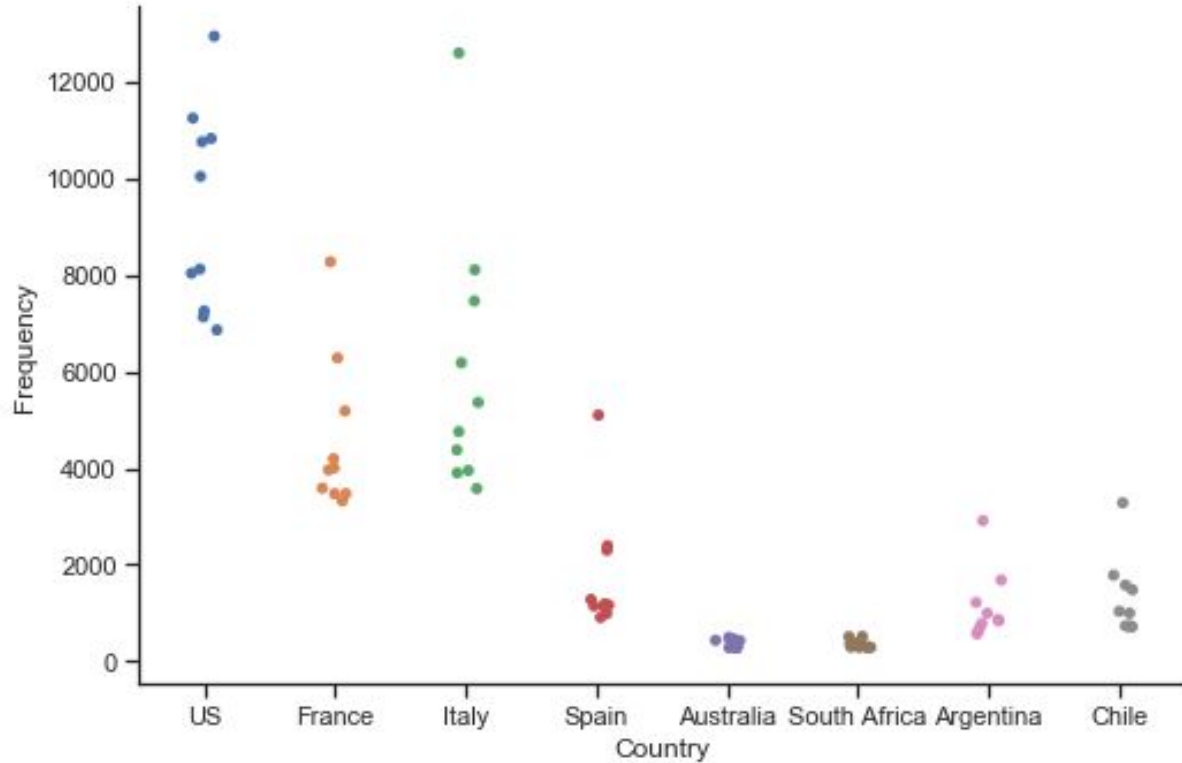


Text Analysis

We compared descriptions of the 8 top wine producing countries in terms of the most frequently appearing descriptive words.

- Overall
- By Groups:
 - Good - 80-84 points
 - Better - 85-89 points
 - Best - 90+

Text Analysis - Difference in Volumes



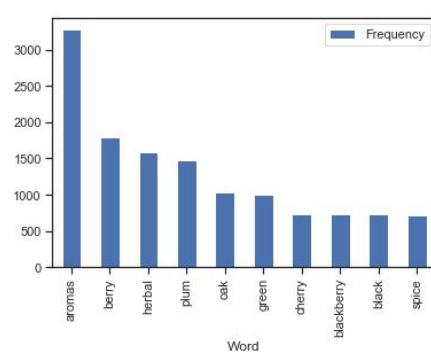
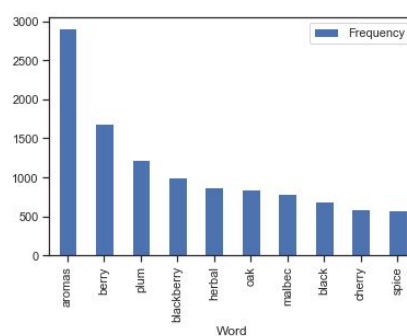
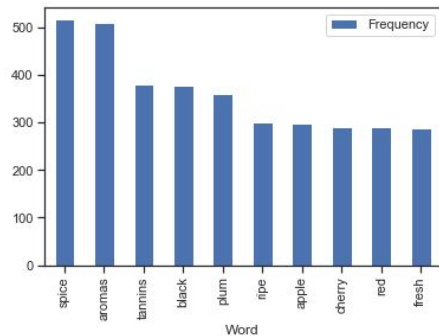
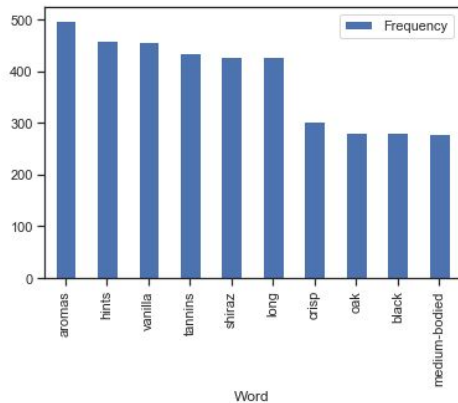
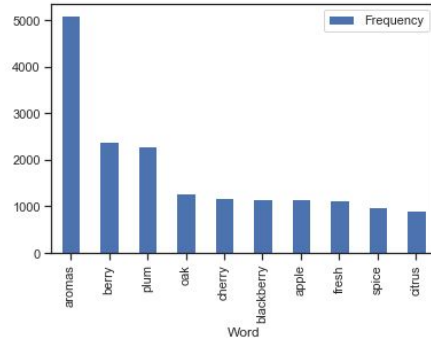
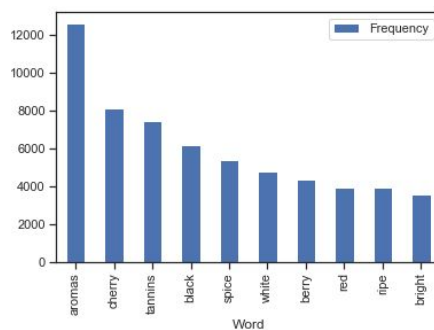
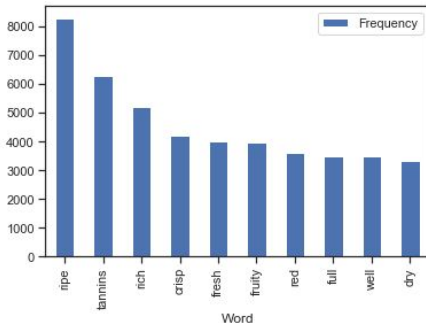
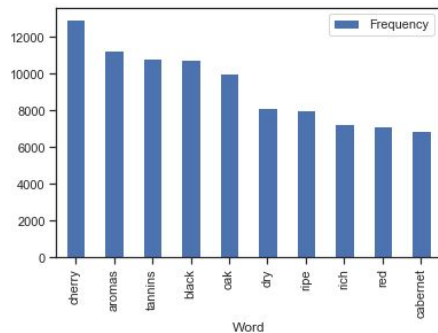
Most Frequent Descriptive Words - Countries

	Word	Frequency	Country
0	cherry	12944	US
1	aromas	11247	US
2	tannins	10826	US
3	black	10757	US
4	oak	10041	US
5	dry	8121	US
6	ripe	8037	US
7	rich	7256	US
8	red	7133	US
9	cabernet	6863	US

	Word	Frequency	Country
0	ripe	8274	France
1	tannins	6283	France
2	rich	5186	France
3	crisp	4197	France
4	fresh	4006	France
5	fruity	3965	France
6	red	3587	France
7	full	3474	France
8	well	3471	France
9	dry	3323	France

	Word	Frequency	Country
0	aromas	12596	Italy
1	cherry	8106	Italy
2	tannins	7462	Italy
3	black	6183	Italy
4	spice	5362	Italy
5	white	4757	Italy
6	berry	4378	Italy
7	red	3952	Italy
8	ripe	3902	Italy
9	bright	3578	Italy

Most Frequent Descriptive Words - Countries



Most Frequent Descriptive Words - Overall

Frequency	
Word	
aromas	36147
tannins	25388
cherry	23857
ripe	20514
black	19013
red	14963
oak	13483
rich	12442
dry	11444
berry	10242

Most Frequent Descriptive Words - Groups

Good Frequency

Word	
aromas	2598
berry	829
red	793
soft	756
light	741
herbal	718
green	706
plum	655
cherry	619
fruity	604

Better Frequency

Word	
aromas	14090
tannins	8826
fresh	7347
ripe	6690
red	5828
berry	5481
cherry	5436
crisp	4798
fruity	4612
black	3960

Best Frequency

Word	
ripe	8626
tannins	7955
aromas	6442
black	5762
rich	5082
cherry	3316
berry	3209
spice	3074
well	2922
full	2856