Introduction

Technical Presentation
"Grape Capital – Wine Investment Thesis"

Analyst: Bart Teeuwen
Date: February 13th, 2021

Agenda

- Problem Summary
- Technical Approach
- Key insights & recommendations
- Wine analysis
    - Wine quality
    - Wine countries
    - Wine ratings
    - Wine varieties
    - Wine price & points
    - Wine climate
    - Wine climate prediction
- Recap
- Winery investments
- Q&A

Problem Summary

1. Grape Capital

- Venture capital firm investing in high growth industries
- Witnessed high growth in wine industry
- Looking to acquire a stake in this industry
- Initial investment of $10M in wineries in next 5 years with 20% ROI

2. Problem

- No previous knowledge of wine industry
  - Wine quality
  - Wineries
  - Wine climate
  - Wine countries

3. Project

- Research wine industry:
  - Wine quality understanding
  - Wine ratings, points, and price foundation
  - Wine varieties & country locations
  - Effect of climate change

- Recommendations for which wineries to invest in based on thorough data analysis

Technical Approach

Elements to explore

Wine quality

Python

- Used Jupyter Notebook to analyze what variables of wine determine quality for red and white wine by performing descriptive and inferential statistical analysis with libraries statsmodels, seaborn, and pandas.
- Visualization tools boxplots, heatmaps, quartiles, barplot, and OLS regression to look at wine variables and how they correlate to wine quality ratings

Excel

- Excel was used for data cleaning and reordering to prep transformation in Tableau

Tableau

Wine reviews
Wine varieties
Wine countries
Wine price
Wine points
Wine climate & prediction

- Used Tableau to analyze wine reviews by taster, points, winery, and country for insights

- Used Tableau to analyze wine varieties on its own and by average price for insights

- Used Tableau to analyze wine varieties on its own, with wine points, with wineries and wine reviews, average price, and climate temperature for insights

- Used Tableau to analyze wine price with wine variety, country, winery, taster, points, and wine reviews for insights

- Used Tableau to analyze wine price with country, taster, wine reviews, and average price for insights

- Used Tableau to analyze average temperature with time in years and countries for insights

Powerpoint

- Powerpoint was used to bring the data together from the Jupyter Notebook and Tableau dashboards

Key Insights & Recommendations

Recommendations

- Split the $10M in two investment groups of short and long-term investments:
  - Short-term – Invest 70% in four wineries based in the US and Portugal between 2021 and 2025
  - Long-term – Invest 30% in three wineries based in Argentina, Spain, and Australia
- Advise wineries of nuances in wine variety preferences for red/white wine to help better market wines
- Consider creating another investment fund if wineries perform above expectation

Key learnings

- Creating high quality red and white wine requires a delicate balance of variables, with density as the main driver
- American wines have more ratings and a lower price points than French wines, which stand out with a high average price
- Wine varieties produced in small regions in a complex climate tends to differ by country and has a higher average price point
- Red wine varieties seemed more popular than white wines based on wine ratings
- Most wineries are in the US followed by France and Italy. Other EU countries are up and coming
- The top 3 winery locations also score most wine review points followed by Mediterranean countries
- A higher wine price does not ensure higher point reviews by (experienced) tasters
- Typically, countries with a Mediterranean climate are a better choice to operate wineries
- Countries not known as a wine country are hosting more wineries, which could prove to become an opportunity eventually
- Europe is well suited for producing wine based on average temperature
- The average climate temperature is set to reach 20 degrees Celsius by 2025, which may provide some countries with a better climate to produce wine while other countries climate could become disadvantageous

Problem

- Grape Capital invests in high growth industries but has observed a large growth in the wine industry
- The company wants a piece of the pie and plans on investing $10M in wineries that it thinks can generate an ROI of 20%
- Grape Capital doesn't have extensive knowledge about the global wine industry and which wineries are successful
- A data analyst is hired to perform data analyses across four datasets: wine quality (1), wineries (2), wine countries (3), and climate change (4)
- The goal is to make recommendation for which wineries Grape Capital should invest its money to achieve it's ROI
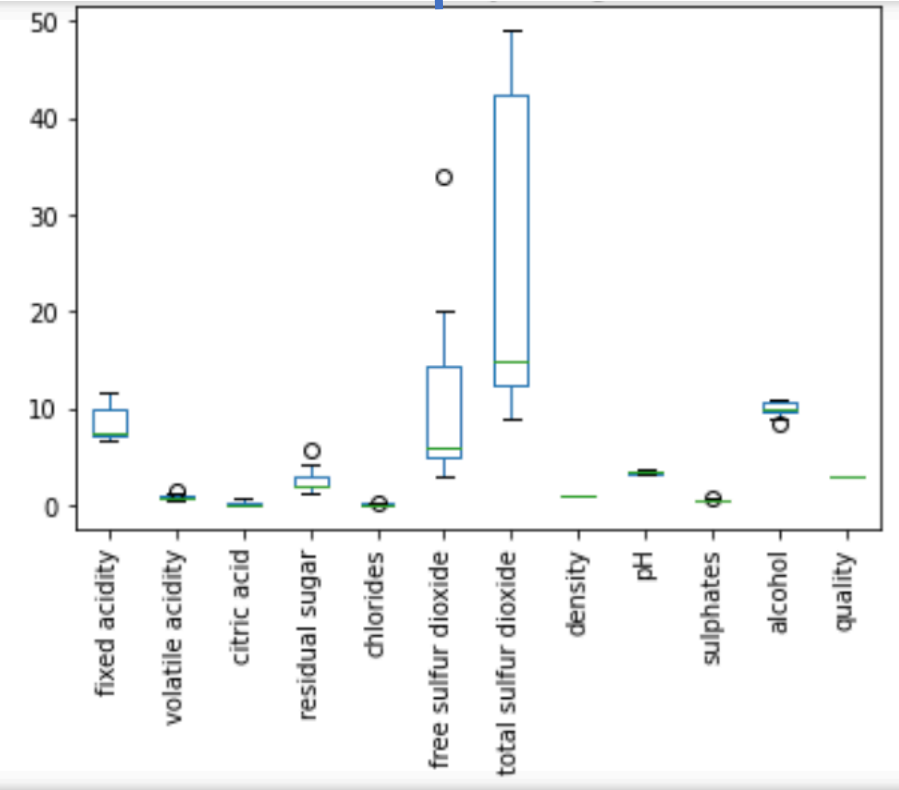
Wine analysis:
Wine quality
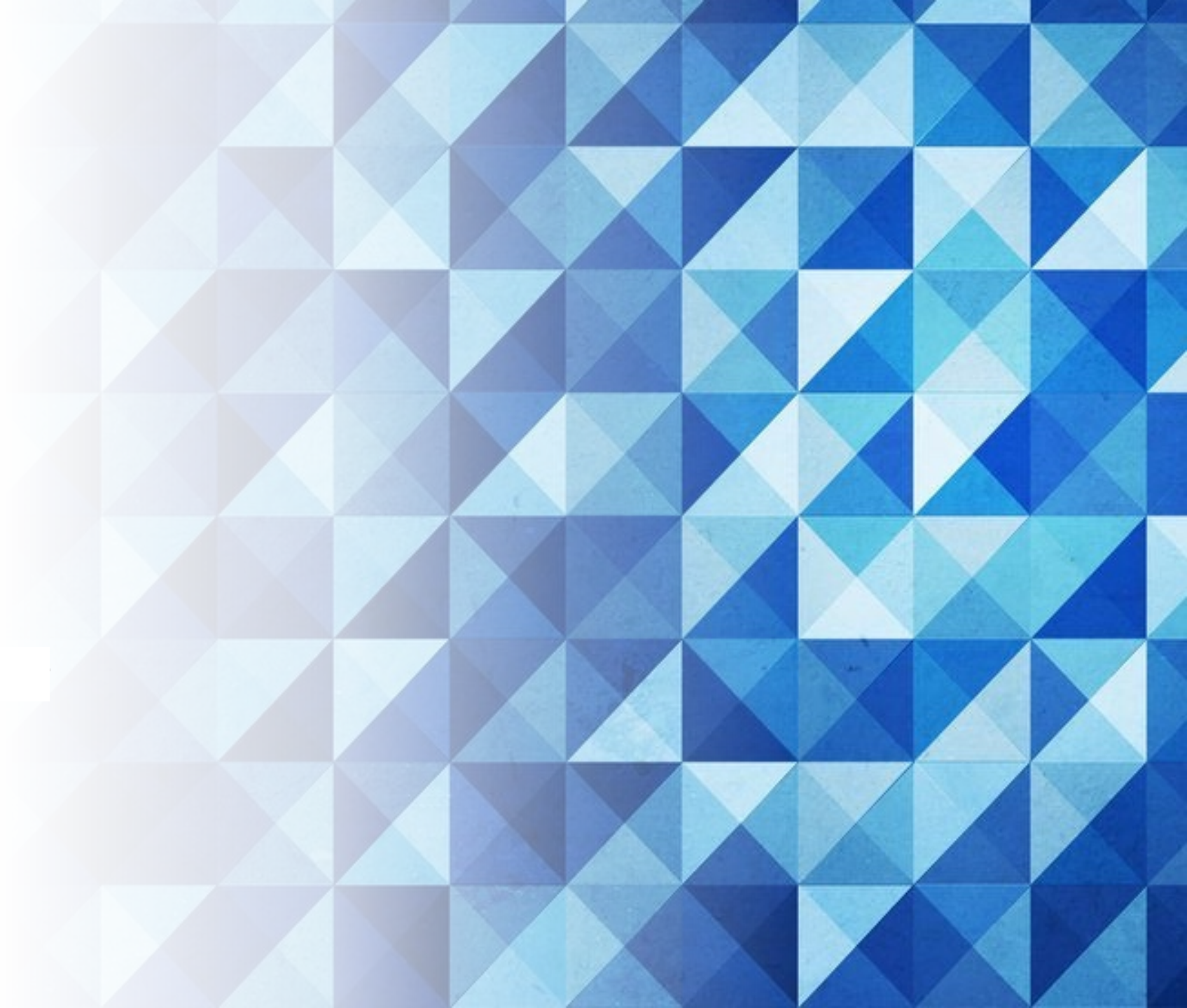
Red Wine Quality – Data suggests that red wine with a higher wine quality rating has between 6-15 and 15-42 values for free and total sulfur dioxide versus lower quality red wine with a wider range

Low Red Wine Quality - 3

- Lower quality red wine seems to have a wider range of values in fixed acidity, free & total sulfur dioxide, and alcohol percentage
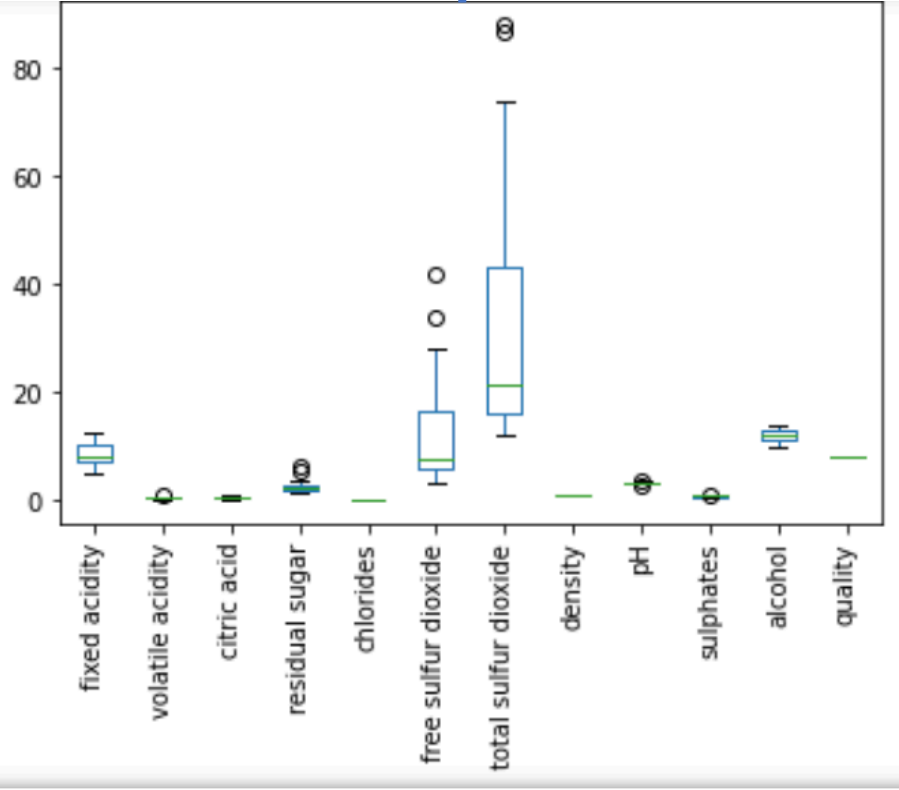
High Red Wine Quality - 8

- Higher quality red wine seems to have a narrow range of values in fixed acidity, free & total sulfur dioxide, and alcohol percentage

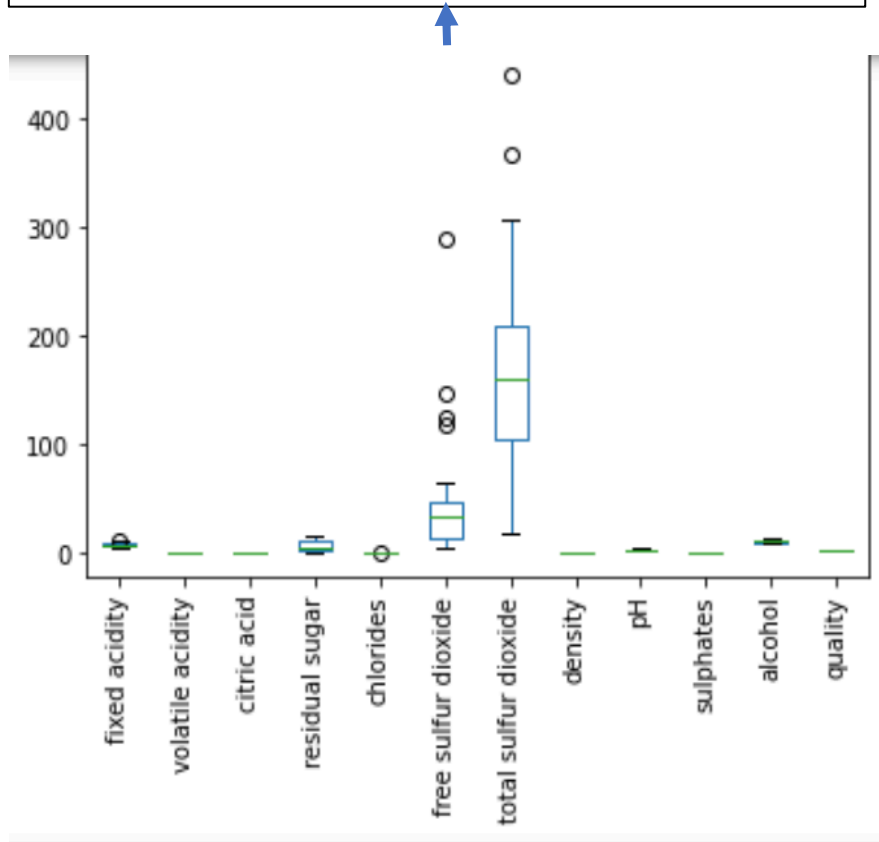*The data for both red and white wine shows there seem to be outliers involved*

* Red wine quality ranges between ratings 3 to 8

White Wine Quality – Data suggests that white wine with a higher wine quality rating has between 22-30 and 115-125 values for free and total sulfur dioxide versus lower quality white wine with a wider range

Low White Wine Quality - 3

- Lower quality white wine seems to have a wider range of values in free & total sulfur dioxide, and a lower alcohol percentage
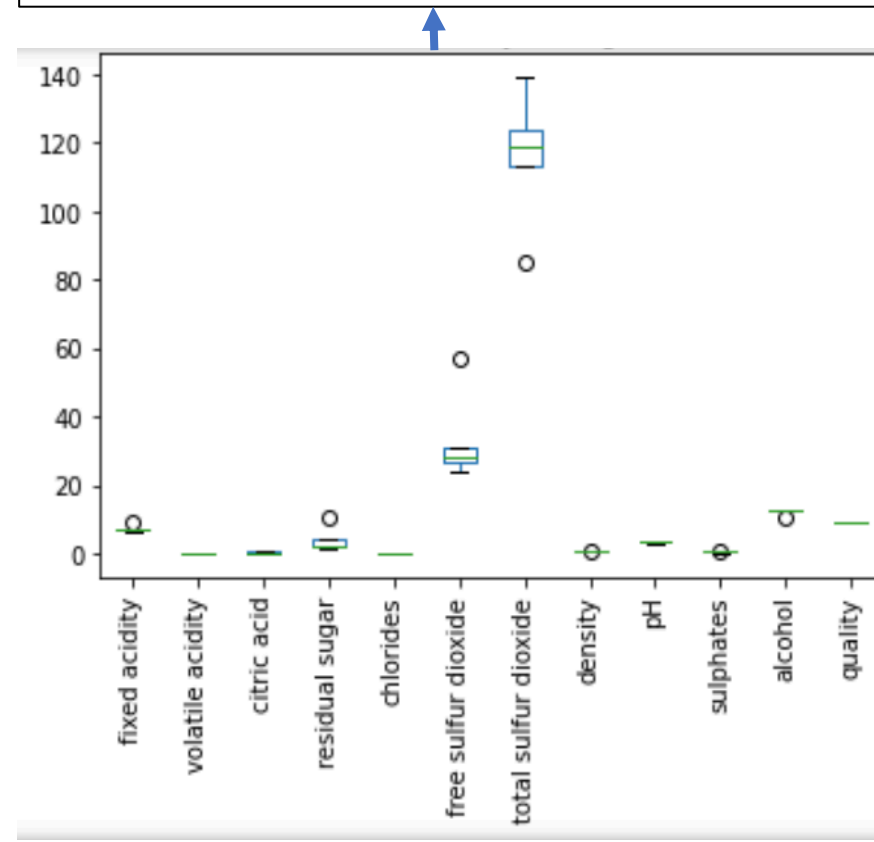

High White Wine Quality - 9

- Higher quality white wine seems to have a narrow range of values in free & total sulfur dioxide, and a higher alcohol percentage

The data for both red and white wine shows there seem to be outliers involved

* White wine quality ranges between ratings 3 to 9

Removing Outliers Red & White Wine – Removing the outliers did not seem to have an adverse affect on the value ranges that point to a higher quality wine rating

High Red Wine Quality - 8

- Removing the outliers (420 – 26%) did not seem to have a major impact on the values that impact higher quality wine
- Main value ranges are:
  - Fixed acidity – 8-10
  - Free sulfur dioxide – 5-15
  - Total sulfur dioxide – 18-42
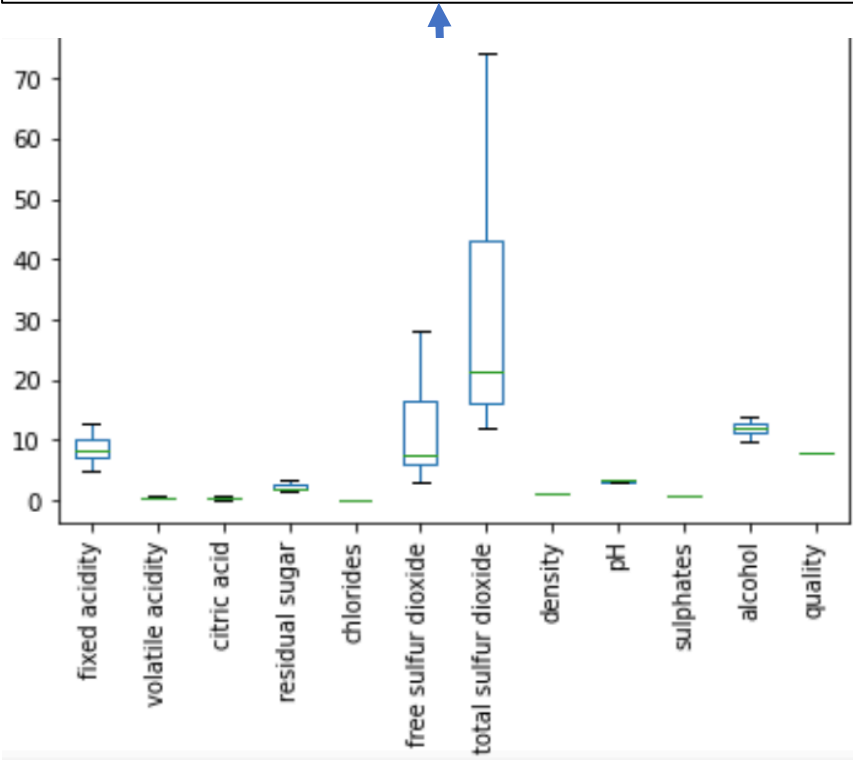
High White Wine Quality - 9

- Removing the outliers (1044 – 21%) did not seem to have a major impact on the values that determine higher quality wine
- Main value ranges are:
  - Free sulfur dioxide – 29-35
  - Total sulfur dioxide – 115-125

Wine Quality Correlation – Heat maps showed that both red wine (pH & fixed acidity) and white wine (alcohol & density) had multilinearity which made it more difficult to determine exactly what wine variables correlate with wine quality

Creating a high-quality wine is a delicate balance of variables where ingredients should not be used either too much or too little
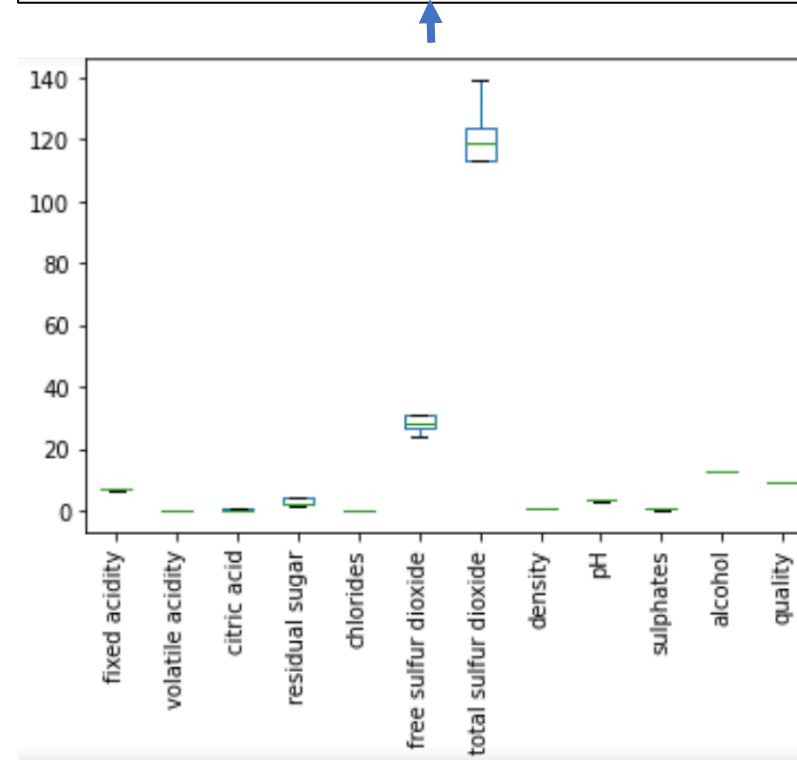
Red Wine Quality Correlation

Red wine correlation:
- Volatile acidity = - 0.39 (medium negative)
- Sulphates = 0.25 (weak positive)
- Alcohol = 0.48 (medium positive)

Medium negative correlation

Correlated to fixed acidity

Weak positive correlation

Medium positive correlation

Heatmap Correlations All Variables for dataframe_red
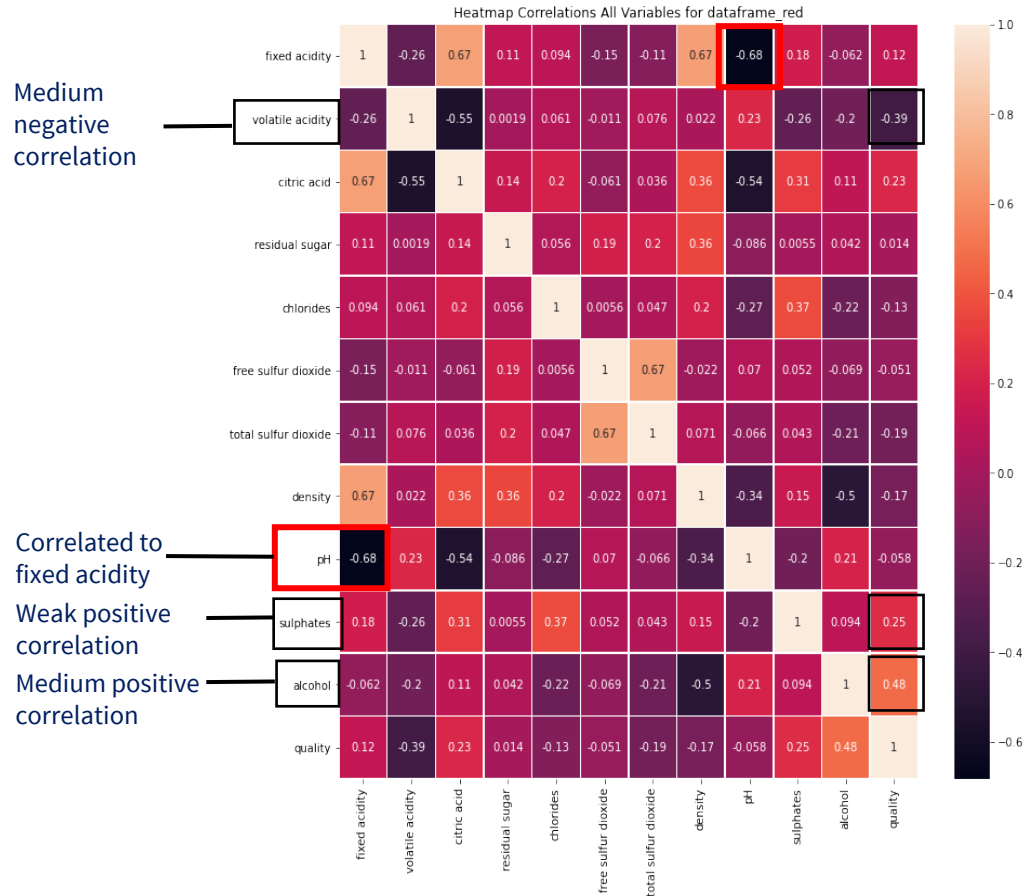
White Wine Quality Correlation

White wine correlation:
- Volatile acidity = - 0.19 (weak negative)
- Density = 0.25 (medium positive)
- Alcohol = 0.48 (medium positive)

Weak negative correlation

Medium negative correlation

Medium positive correlation

Heatmap Correlations All Variables for dataframe_white

Correlated to density
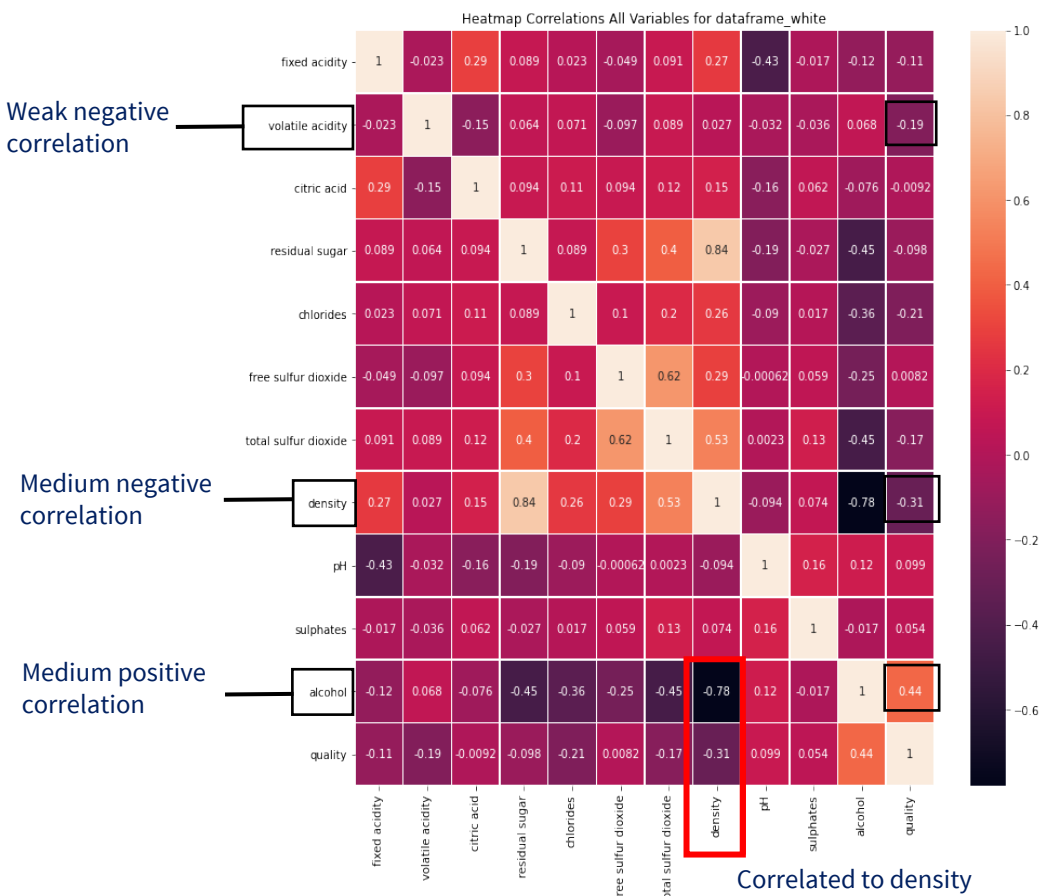
OLS Regression – Removing insignificant variables (P- value >0.05) for both red and white wine singled out chlorides and density for having the largest impact on wine quality

Red Wine:
Significant wine variables

- Volatile acidity, chlorides, and pH are negatively correlated to wine quality (more of them results in lower wine quality)
- Density, sulphates, and alcohol are positively correlated to wine quality (more of them results in higher wine quality)

Dep. Variable:	quality	R-squared (uncentered):	0.987
Model:	OLS	Adj. R-squared (uncentered):	0.987
Method:	Least Squares	F-statistic:	1.736e+04
Date:	Sun, 28 Feb 2021	Prob (F-statistic):	0.00
Time:	18:24:13	Log-Likelihood:	-1573.6
No. Observations:	1599	AIC:	3161.
Df Residuals:	1592	BIC:	3199.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
volatile acidity	-1.0395	0.100	-10.350	0.000	-1.237	-0.843
chlorides	-1.9948	0.398	-5.012	0.000	-2.775	-1.214
total sulfur dioxide	-0.0024	0.001	-4.657	0.000	-0.003	-0.001
density	4.2274	0.394	10.730	0.000	3.455	5.000
pH	-0.4207	0.115	-3.656	0.000	-0.646	-0.195
sulphates	0.8813	0.110	7.999	0.000	0.665	1.097
alcohol	0.2943	0.017	17.593	0.000	0.262	0.327

Medium negative correlation

Strong positive correlation

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[3] The condition number is large, 2.47e+03. This might indicate that there are strong multicollinearity or other numerical problems.

White Wine:
Significant wine variables

- Chlorides and volatile acidity are negatively correlated to wine quality (more of them results in lower wine quality)
- Density, sulphates, and pH are positively correlated to wine quality (more of them results in higher wine quality)

Dep. Variable:	quality	R-squared (uncentered):	0.980
Model:	OLS	Adj. R-squared (uncentered):	0.980
Method:	Least Squares	F-statistic:	3.034e+04
Date:	Sun, 28 Feb 2021	Prob (F-statistic):	0.00
Time:	15:57:56	Log-Likelihood:	-6068.6
No. Observations:	4898	AIC:	1.215e+04
Df Residuals:	4890	BIC:	1.221e+04
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
fixed acidity	-0.0638	0.016	-3.952	0.000	-0.095	-0.032
volatile acidity	-1.2791	0.122	-10.487	0.000	-1.518	-1.040
chlorides	-6.7818	0.563	-12.051	0.000	-7.885	-5.679
free sulfur dioxide	0.0072	0.001	7.824	0.000	0.005	0.009
total sulfur dioxide	-0.0047	0.000	-12.493	0.000	-0.005	-0.004
density	6.0856	0.355	17.159	0.000	5.390	6.781
pH	0.3402	0.089	3.810	0.000	0.165	0.515
sulphates	0.4987	0.107	4.651	0.000	0.288	0.709

Medium negative correlation

Strong negative correlation

Strong positive correlation

Wine Quality Conclusion

1. The OLS regression model predicted wine quality values with 98% confidence
2. Both red wine and white wine variables showed multilinearity which could skew significance toward wine quality
3. Removing multilinearity variables fixed acidity for red wine, alcohol for white wine, and other non-significance variables created a stronger and clearer model for estimating how changes in the independent wine variables are associated with changes in the dependent wine quality variable
4. Both red and white wine regression data suggests density had a positive coef while volatile acidity and chlorides had a negative coef to wine quality

Wine Countries – The United States, France, and Italy host most wineries and points globally

Wine Countries

- The United States, France, and Italy are the top 3 winery destinations
- The US has more than double the number of wineries France has

Countries with Most Wineries

Country1	
United States	54,504
France	22,093
Italy	19,540
Spain	6,645
Portugal	5,691
Chile	4,472
Argentina	3,800
Austria	3,345
Australia	2,329
Germany	2,165
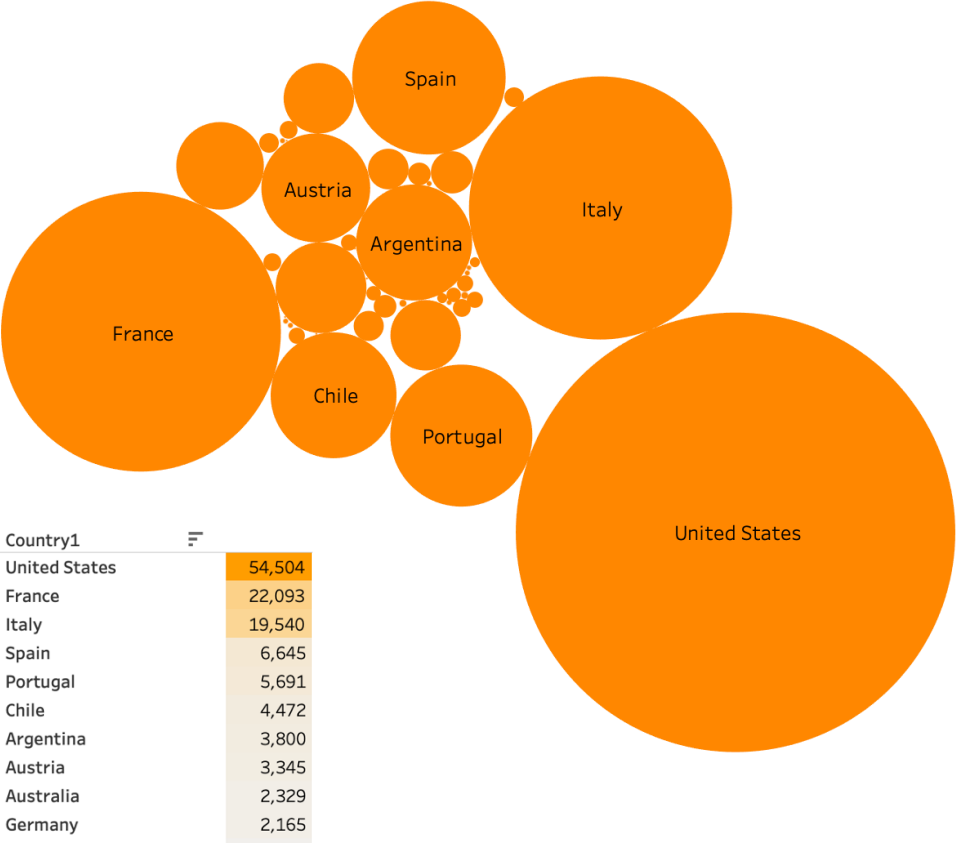
Wine Points

- The United States, France, and Italy scored the most wine points
- Mediterranean countries seem to pick up in points
- The US has almost 2.5X more points than France

Countries with most Points

Country1	
United States	4,827,077
France	1,962,855
Italy	1,730,506
Spain	580,031
Portugal	502,232
Chile	386,799
Argentina	329,499
Austria	301,389
Australia	206,304
Germany	194,529
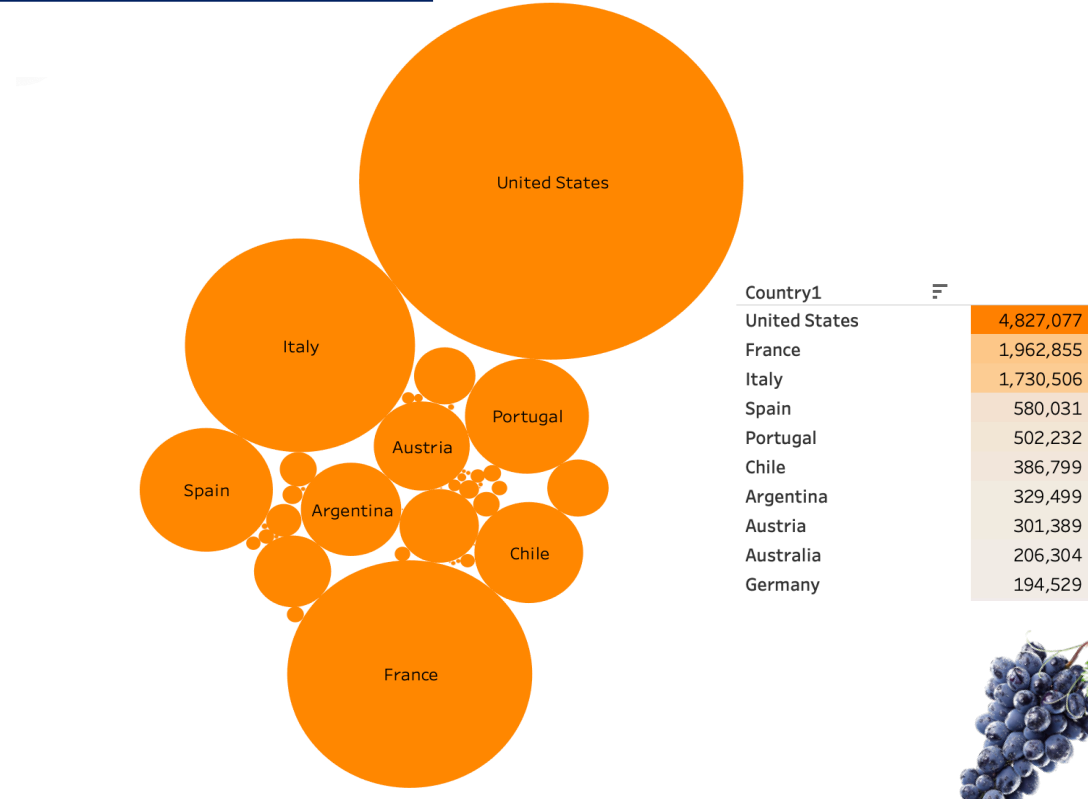
Wine Ratings – The top 15 wineries all have 100+ reviews but cost less than $135 avg while the top 15 most expensive wineries have <30 reviews

Winery Reviews

- The United States has a lot of wineries that have a lot of reviews compared to other countries
- France still maintains its presence as a large wine country

Winery avg Prices

- 93% of the top 15 most expensive wineries by avg price is France
- Each of the top 15 wineries has less than 30 reviews overall
- Wineries with the most reviews cost on avg cost less than $135 a bottle

Wineries with Most Reviews

Winery1	Country1	
Wines & Winemakers	Portugal	222
Testarossa	United States	218
DFJ Vinhos	Portugal	215
Williams Selyem	United States	211
Louis Latour	France	199
Georges Duboeuf	France	196
Chateau Ste. Michelle	United States	194
Concha y Toro	Chile	164
Columbia Crest	United States	159
Kendall-Jackson	United States	130
Siduri	United States	126
Gary Farrell	United States	125
Lynmar	United States	118
Montes	Chile	117
Albert Bichot	France	117

Wineries with Highest AVG Price

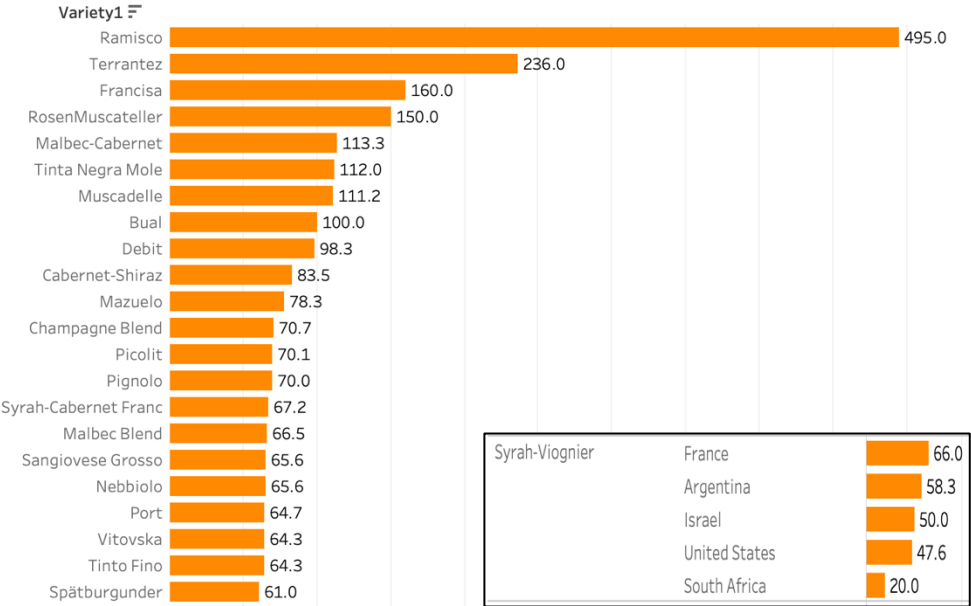| Winery1	Country1	
Château les Ormes Sorbet	France	3,300
Château Pétrus	France	2,250
Domaine du Comte Liger-Belair	France	1,489
Château Cheval Blanc	France	825
Château d'Yquem	France	800
Masseto	Italy	588
Château Laville Haut-Brion	France	580
Château Haut-Brion	France	572
Château La Mission Haut-Brion	France	546
Château d'Ausone	France	507
Château Mouton Rothschild	France	479
Château Lafite Rothschild	France	473
Armand de Brignac	France	469
Château Margaux	France	448
Château Latour	France	437

Wine Varieties – The most popular wine varieties are red wines, and there seems to be a large price difference by wine variety due to scarcity

Wine Varieties & Avg Price

- Ramisco, Terrantez, and RosenMuscateller have the highest avg price (rare Portugese/Australian grapes)
- Wine varieties range in price by geography

Wine Varieties with High Price

Variety1	
Ramisco	495.0
Terrantez	236.0
Francisa	160.0
RosenMuscateller	150.0
Malbec-Cabernet	113.3
Tinta Negra Mole	112.0
Muscadelle	111.2
Bual	100.0
Debit	98.3
Cabernet-Shiraz	83.5
Mazuelo	78.3
Champagne Blend	70.7
Picolit	70.1
Pignolo	70.0
Syrah-Cabernet Franc	67.2
Malbec Blend	66.5
Sangiovese Grosso	65.6
Nebbiolo	65.6
Port	64.7
Vitovska	64.3
Tinto Fino	64.3
Spätburgunder	61.0

Syrah-Viognier		
	France	66.0
	Argentina	58.3
	Israel	50.0
	United States	47.6
	South Africa	20.0

Wine Varieties

- Pinot Noir, Chardonnay, Cabernet Sauvignon, Red Blend, and Bordeaux-Style Red Blends are the popular wine varieties
- Red wine varieties seem to be more popular than white wine varieties

Most Popular Wine Varieties

Variety1	
Pinot Noir	13,272
Chardonnay	11,753
Cabernet Sauvignon	9,472
Red Blend	8,946
Bordeaux-style Red Blend	6,915
Riesling	5,189
Sauvignon Blanc	4,968
Syrah	4,142
Rosè	3,564
Merlot	3,102
Nebbiolo	2,804
Zinfandel	2,714
Sangiovese	2,707
Malbec	2,652
Portuguese Red	2,466

Wine Price & Points Comparison – A higher wine price does not ensure higher point reviews by (experienced) tasters

Rating Count & Points by Taster

- Roger Voss, Michael Schachner, and Kerin O, ÄôKeefe gave out the most wine ratings from all tasters
- Roger Voss gave 1.7X more wine ratings than the 2nd closest taster in terms of wine ratings

Wine Rating Count and Points by Taster

Taster Name	Count of Wine ratings	Points
Roger Voss	25,514	2,263,296
Michael Schachner	15,134	1,315,258
Kerin O,ÄôKeefe	10,776	957,641
Virginie Boone	9,537	850,828
Paul Gregutt	9,532	849,135
Matt Kettmann	6,332	569,935
Joe Czerwinski	5,147	455,696
Sean P. Sullivan	4,966	440,761
Anna Lee C. Iijima	4,415	390,355
Jim Gordon	4,177	370,192
Anne Krebiehl¬†MW	3,685	333,723
Lauren Buzzeo	1,835	161,002
SUnited Statesan Kostrze..	1,085	93,971
Mike DeSimone	514	45,798
Jeff Jenssen	491	43,365
Alexander Peartree	415	35,630
Carrie Dykes	139	12,009
Fiona Adams	27	2,346
Christina Pickard	6	527

Comparing avg Rating Points With avg Price & Reviews

- Price does not seem to correlate to more review points
- A large percentage (50% of total) of wines receives a score of at least 86 points
- On average, Anne Krebiehl has given the most points for wine reviews while Virginie Boone review most expensive wines on average
- Michael Schachner and Roger Voss have reviewed most wines from all tasters

Average Points & Price per Wine Review by Taster

| Taster Name	Avg. Points	Avg. Price	Count ..	Percentile (5..
Roger Voss	89	39	25,514	88
Michael Schachner	87	25	15,134	87
Kerin O,ÄôKeefe	89	42	10,776	89
Virginie Boone	89	47	9,537	90
Paul Gregutt	89	34	9,532	89
Matt Kettmann	90	39	6,332	90
Joe Czerwinski	89	35	5,147	89
Sean P. Sullivan	89	34	4,966	89
Anna Lee C. Iijima	88	30	4,415	88
Jim Gordon	89	27	4,177	89
Anne Krebiehl¬†MW	91	31	3,685	90
Lauren Buzzeo	88	24	1,835	88
SUnited Statesan Kostrze..	87	23	1,085	87
Mike DeSimone	89	28	514	89
Jeff Jenssen	88	22	491	89
Alexander Peartree	86	29	415	86
Carrie Dykes	86	31	139	87
Fiona Adams	87	31	27	87
Christina Pickard	88	29	6	88

Wine Climate – Countries in Africa and South America have a better wine climate despite more wines coming from moderate climates

Country Climate

- The AVG temperature differs by country but follows the general equator climate zones

Countries and AVG Temperature
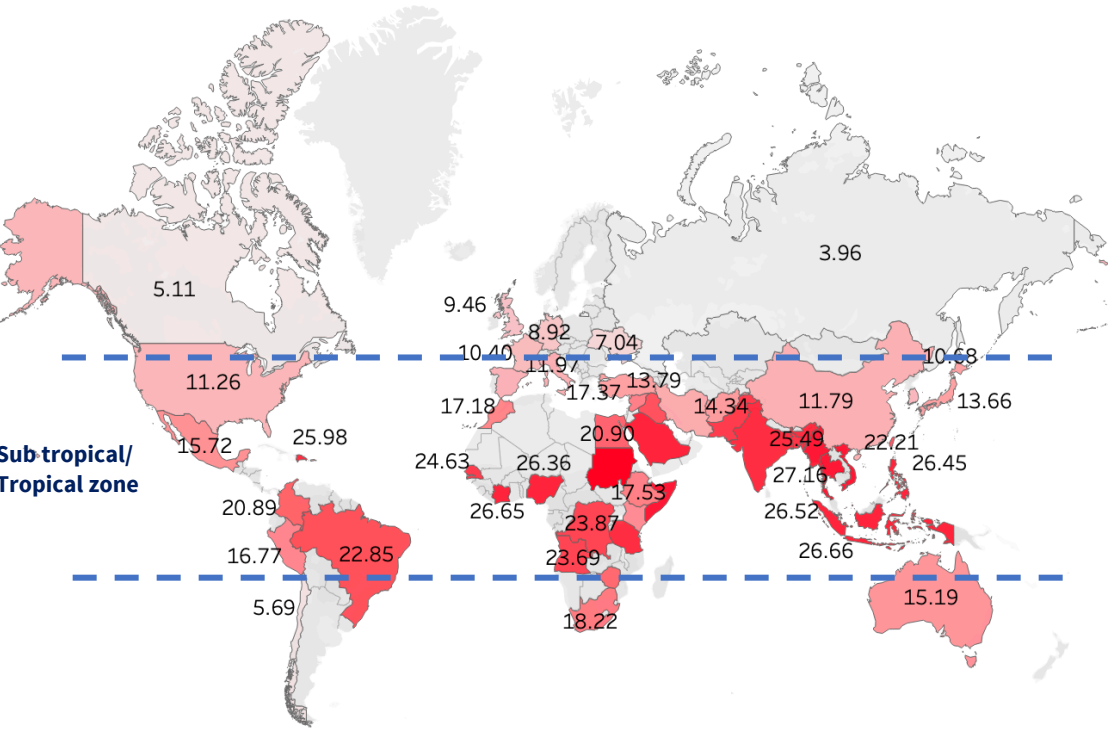
Country Climate Filter

- Filter list of countries with AVG temperatures to match countries from the wine ratings dataset

Country Climate Filter

Ideal Country Climate

- India, Brazil, and Egypt are the top 3 countries based on AVG temperature
- Europe has most countries with ideal climate

AVG Temperature by Country

Country1		Continents		Rank	Score
India	25.81	Asia		Europe	7
Brazil	22.85	South America		South America	3
Egypt	20.90	Africa		Africa	3
South Africa	18.22	Africa		North America	3
Morocco	17.18	Africa		Asia	2
Peru	16.77	South America		Oceania	1
Mexico	15.72	North America			
Australia	15.19	Oceania			
Turkey	13.79	Europe			
Italy	11.97	Europe			
China	11.79	Asia			
Spain	11.45	Europe			
United States	11.26	North America			
France	10.40	Europe			
United Kingdom	9.46	Europe			
Germany	8.92	Europe			
Ukraine	7.04	Europe			
Chile	5.69	South America			
Canada	5.11	North America			

The ideal wine climate is between 10 and 27 degrees Celsius

Wine climate prediction – Globally the average temperature is set to rise 0.15 degrees Celsius by 2025
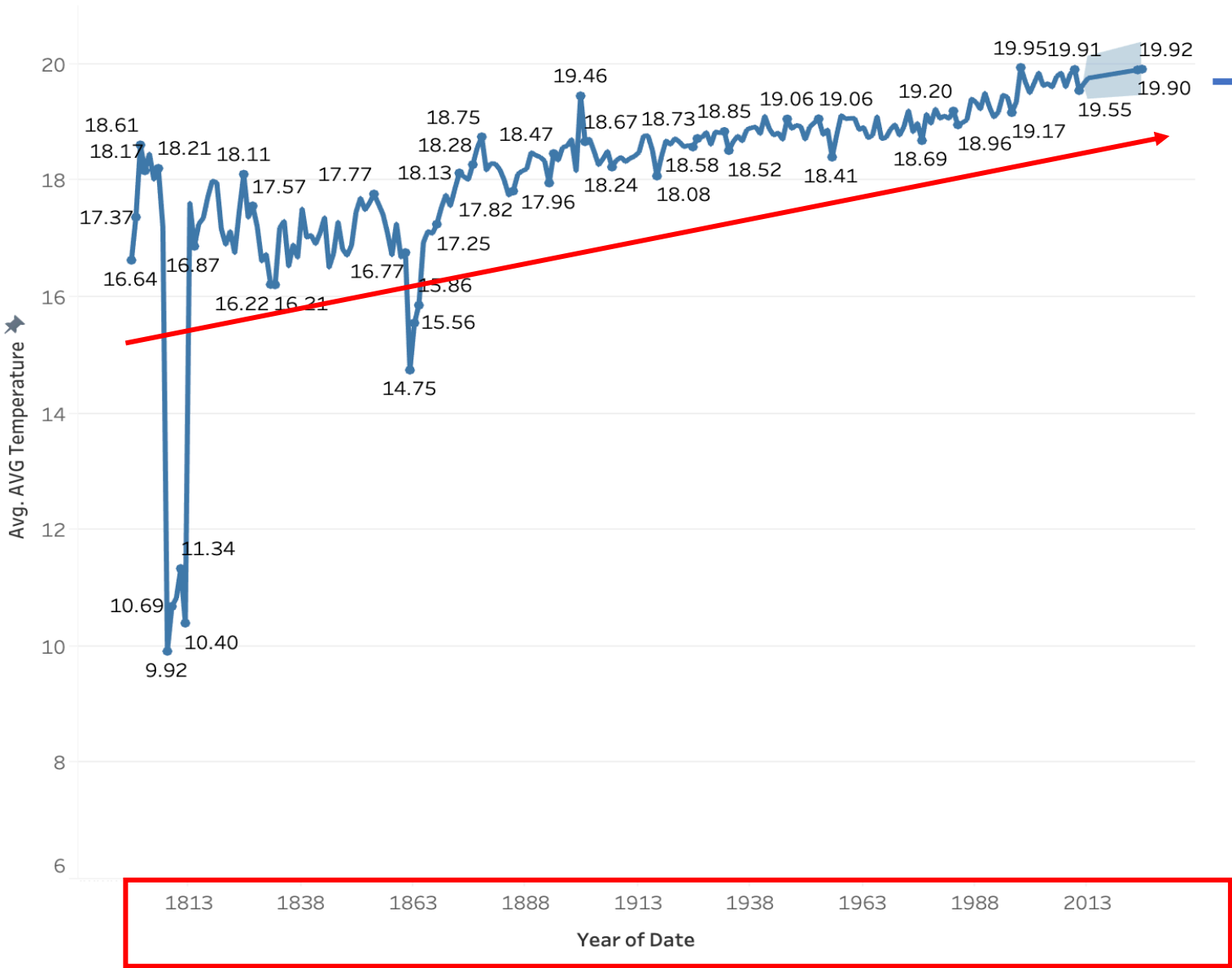
Country Climate Prediction

Avg Temperature Prediction

- Avg temperature is set to rise by 0.15 degrees Celsius between 2013-2025
- Not all countries experience the same change in avg temperature due to climate change, economic policies, or other factors

Country Climate Effect

- Countries with historically lower avg temperature could suddenly bloom into a wine region
- Countries with increasing avg temperature due to climate change, economic policies, or other factors could negatively impact wine production

Recap

So far, we've discussed the following:

- *Wine Capital problem.*
- *Technical approach*
- *Key insights & recommendations*
- *Wine analysis*

> Investment recommendations

Winery Investments – Based on our data analysis of wine variables, countries, ratings, varieties, prices & points, and climate, Grape Capital should consider investing its money in 7 wineries across 5 countries that produce 12 wine varieties

	Country	Winery	Variety	Investment
Short Term 2021-2025	United States	Testarossa Williams Selyem Chateau Ste. Michelle	Pinot Noir Chardonnay Zinfandel White Riesling	$6M
	Portugal	Wines & Winemakers	Portugese Red Touriga Nacional	$1M
Long Term 2025-2030	Argentina	Trapiche (AR)	Shira- Viognier Malbec	$1M
	Spain	CVNE (ES)	Tempranillo Tempranillo Blend	$1M
	Australia	D' Arenberg (AUS)	Shiraz Riesling	$1M
ROI	**5 countries**	**7 wineries**	**12 varieties**	**$200K** (20% * $10M)

Q&A

Image Sources

Slide 2, 6, and 20: https://www.markettraders.com/blog/how-to-trade-triangle-chart-patterns/

All slides' grapes: https://www.clipartkey.com/view/ohRToT_grape-transparent-wine-cabernet-sauvignon-wine-grapes/

Dataset Limitations & Assumptions

Dataset 1 – Wine Quality for red and white wine:

This datasets is related to red and white variants of the Portuguese "Vinho Verde" wine and is part of the UCI machine learning repository (UCI, 2015). The data was collected from May 2004 to February 2007, included 4898 wine samples for each wine variant, and was evaluated by a minimum of three sensory assessors through blind tastings. Each tasting was given a score on a scale that ranges from 0 to 10, that matches to very bad to excellent quality, respectively (white wine scores ranged between 3 and 9).

Dataset 2 – Wine ratings:

The dataset is pretty large with 150K rows and contains 10 columns of wine reviews scraped from WineEnthusiast during June of 2017. Each record in the dataset represents a single wine review from an online user of Wine Enthusiast Magazine.

Dataset 3 – Climate temperature

This dataset was put together by Berkeley Earth for their Berkeley Earth Surface Temperature Study. It combines 1.6 billion temperature reports from 16 pre-existing archives.

Assumptions

- Wine quality data was assumed to be general for the entire white and red wines available
- Wine reviews from 2017 are assumed to have stayed consistent in quality as to warrant the investment recommendations
- Climate temperature data from countries did not match most of the countries listed in the wine rating dataset. The wine rating dataset was chosen as the main dataset feeding into the recommendations
- The climate temperature dataset with bigger list of countries was primarily used to showcase and predict the average temperature changes over time and matched to countries from wine ratings dataset
- ROI is a projection and may not be generated by the projected date as mentioned