

Wine Market Insights



Abhinandan Garg
Jessica Young
Rachel Spreng
Priya Jindal



Agenda

- Project Objectives
- Data Overview & Preprocessing
- Data Explorations
- Developing Business Understanding
- Conclusions



SPARTANS WILL.

Project Objective

Project Objective

- Approached by a new wine producer and distributor
- Provide insight into the wine market using a dataset with reviews from sommeliers
- Goal is to provide the customer effective insight for them to confidently produce highly rated wines



SPARTANS WILL.

Data Overview & Preprocessing

Data Overview

Wine Data :

- **Source:** Kaggle
- **Description:** Wine Data contains 10 columns of 130 thousand reviews
- **Variables:** combination of continuous (price, points), text (review, wine description) and categorical (countries, regions)

Global Temperature Dataset:

- **Source:** Kaggle
- **Description:** Dataset showing world temperatures organized by city, state, country for all major countries spanning over 200 years
- **Variables:** combination of continuous (temperature) and discrete (month, year)

Data Overview – Example Data

- Wine Reviews
- Temperature Database



Preprocessing

- Retained only 24 countries with more than 70 reviews
- Fetched production year from the wine title by using text mining
 - Example: Terre di Giurfo 2013 Belsito Frappato (Vittoria)
- Removed any data with missing reviews, price or points

Preprocessing

- Created a Price Tag to characterize wines by price segment
- Used the wine description to understand the most frequently used words to designate a wine. Visualized this by making a word cloud in R.
 - Example: “Aromas include tropical fruit, broom, brimstone and dried herb. The palate isn’t overly expressive, offering unripened apple, citrus and dried sage alongside brisk acidity”

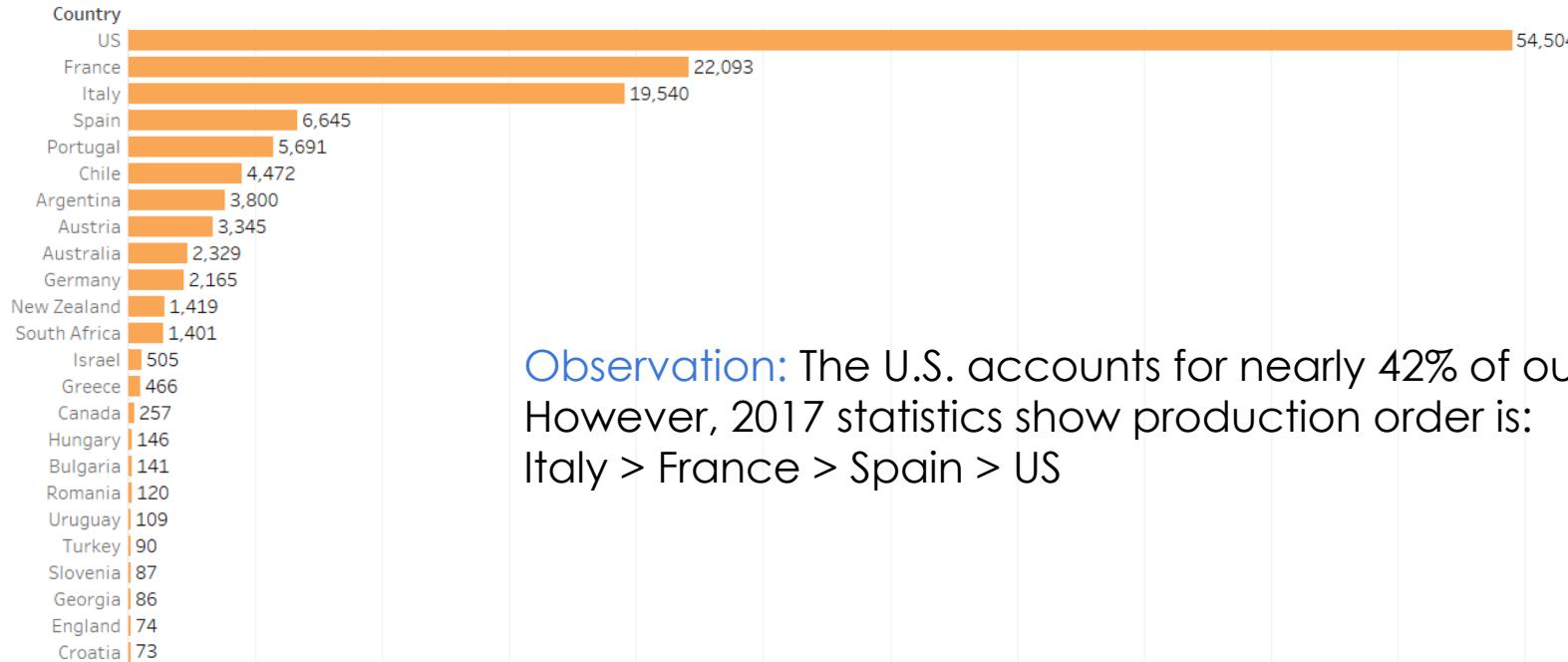


SPARTANS WILL.

Exploratory Data Analysis

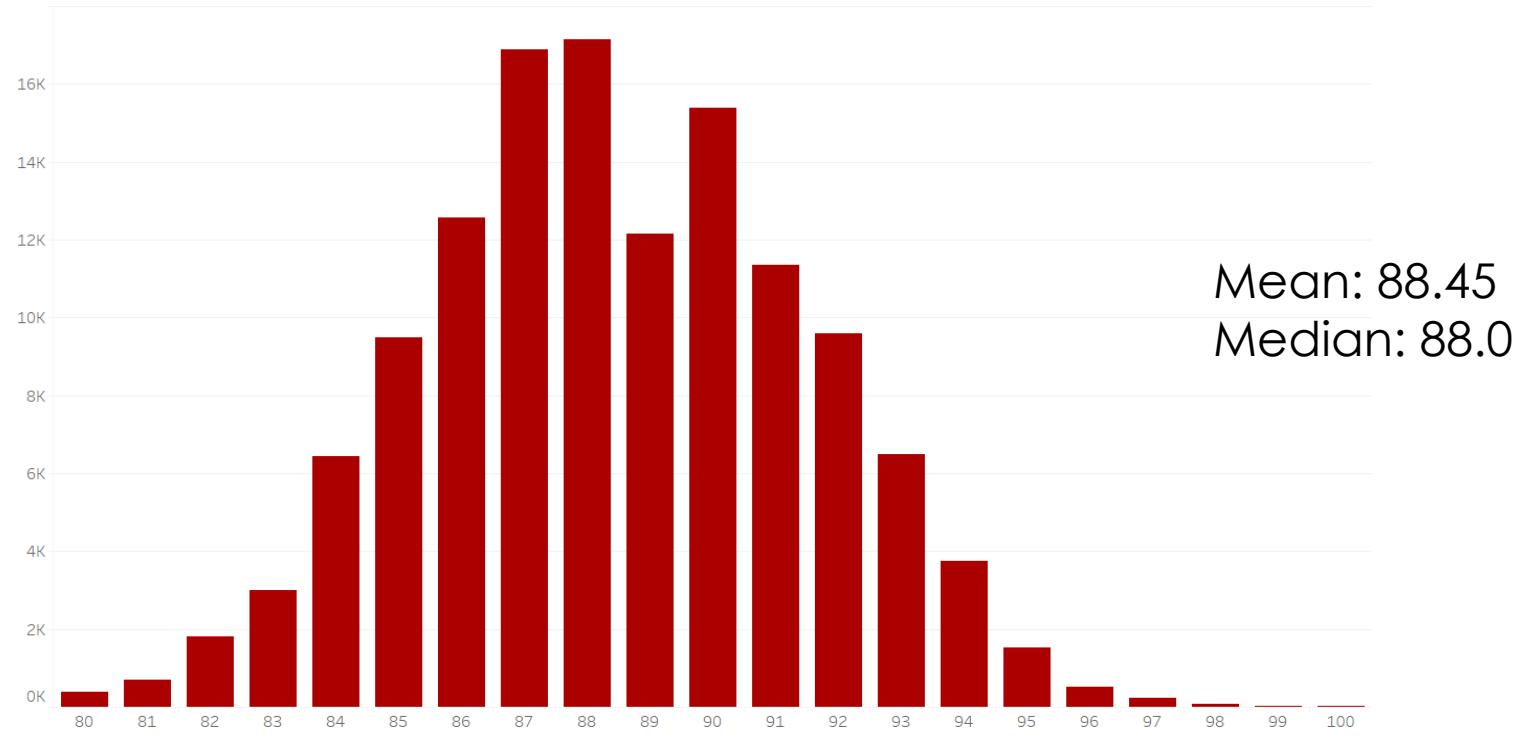
Distribution of Reviews

Number of Reviews by Country

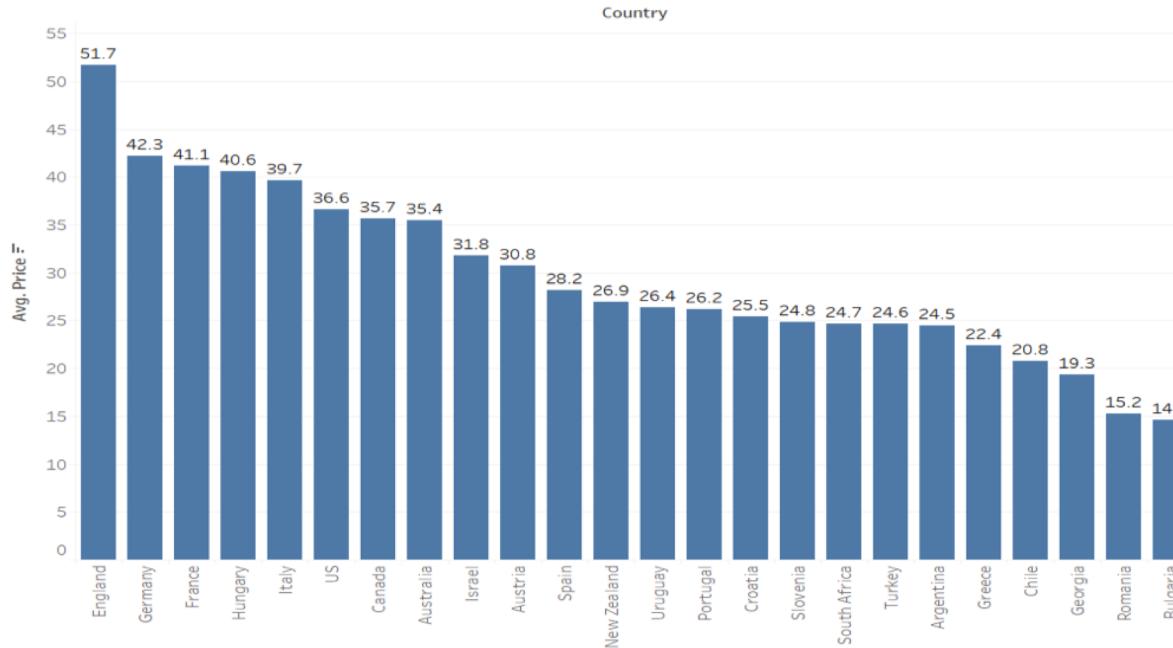


Observation: The U.S. accounts for nearly 42% of our reviews. However, 2017 statistics show production order is:
Italy > France > Spain > US

Distribution of Points

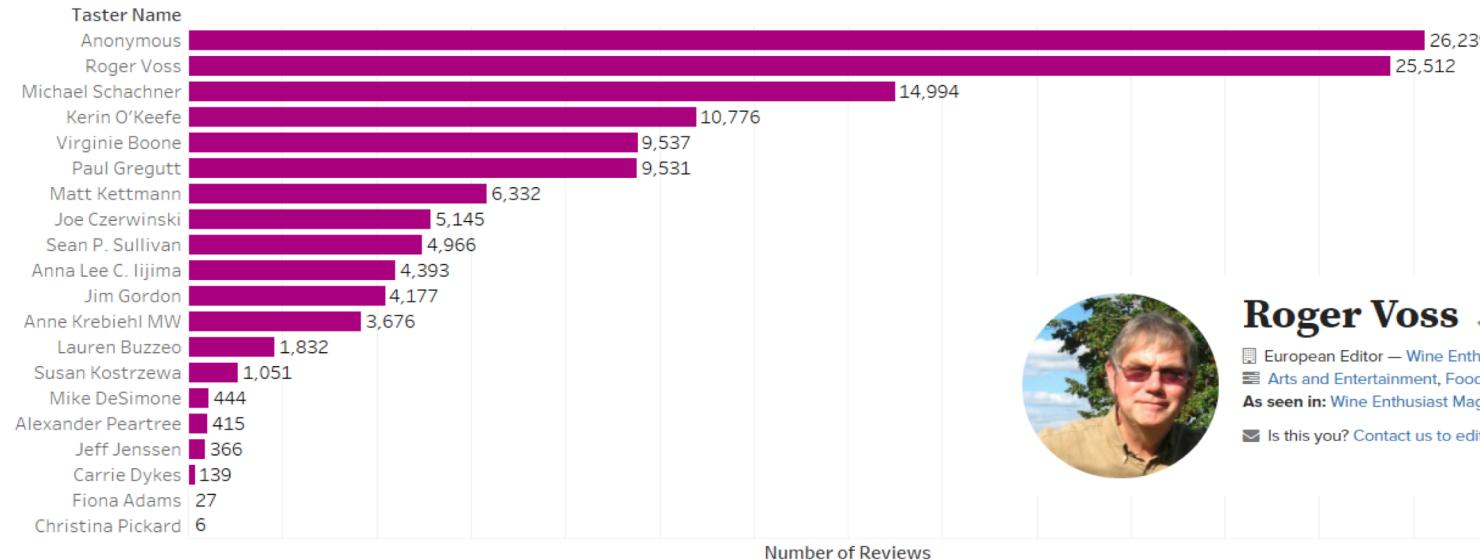


Average Price per Country



Who has reviewed these wines?

Number of Reviews by Taster



Roger Voss 

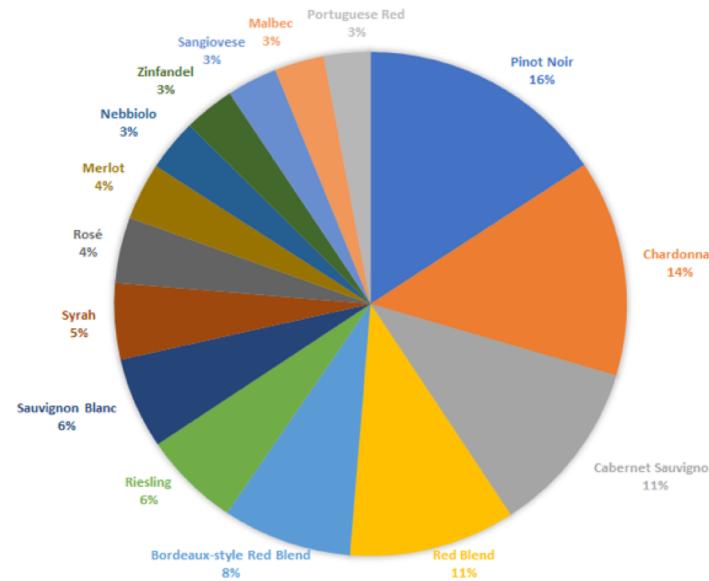
 European Editor — Wine Enthusiast Magazine
 Arts and Entertainment, Food and Dining, France, World
 As seen in: Wine Enthusiast Magazine
 Is this you? Contact us to edit this page

Observation: All reviews were done by sommeliers. A large portion of reviews are coming from a single taster, Roger Voss

Qualifying the reviewers

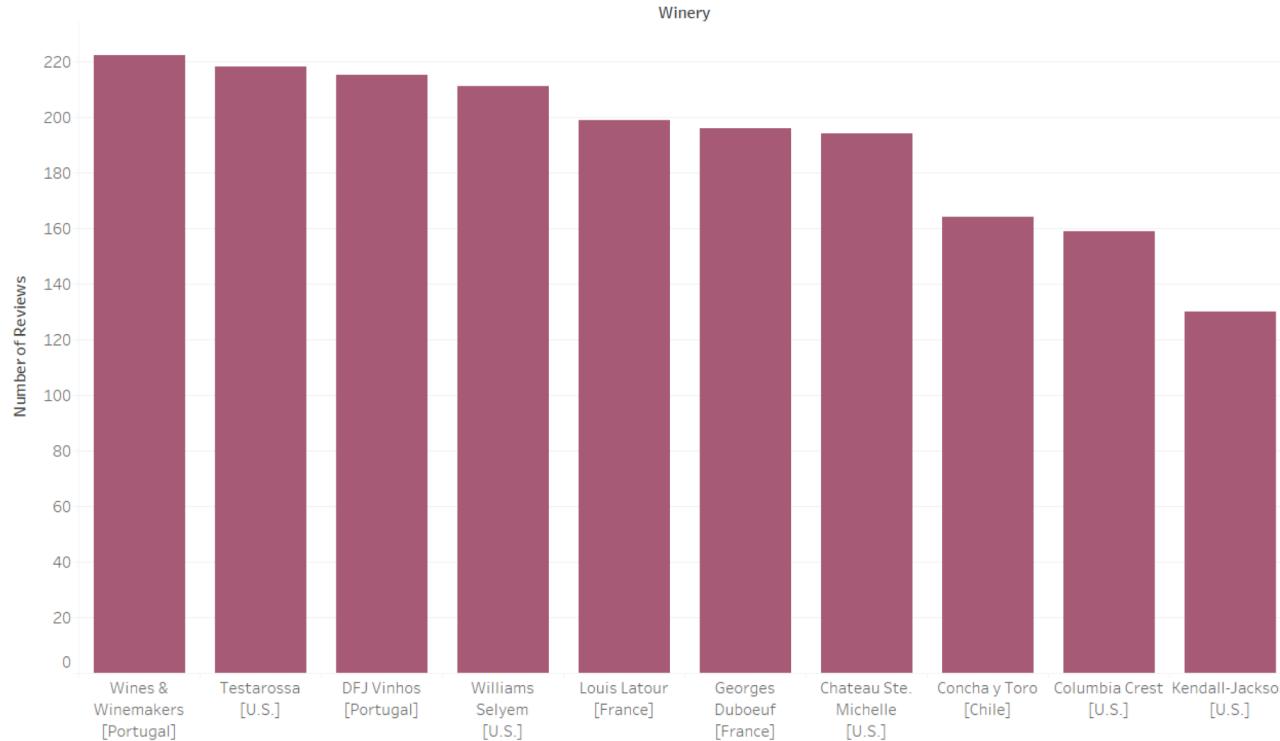
Name	Number of Countries	Total Number of Varieties
Kerin Okeefe	1	140
Roger Voss	7	216
Michael Schachner	8	200

Which wines are most popular?



Observation: Pinot Noir, Chardonnay, Cabernet Sauvignon, and Red Blend are by far the most popular types of wine reviewed by our sommeliers

Which wineries are reviewed the most



Observation: While Portugal only received about 4% of the reviews, two of the top three rated Wineries are located there.



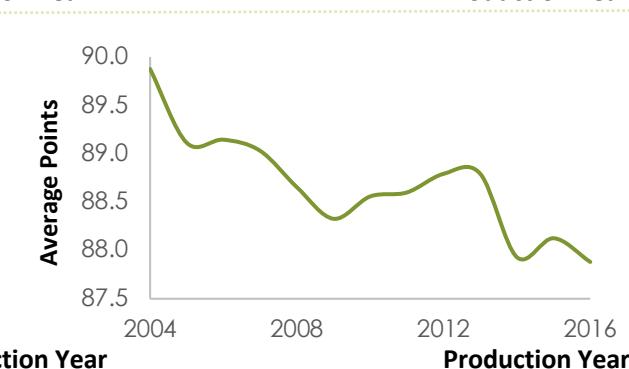
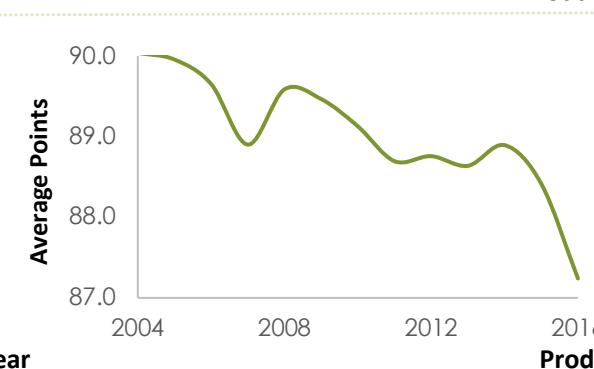
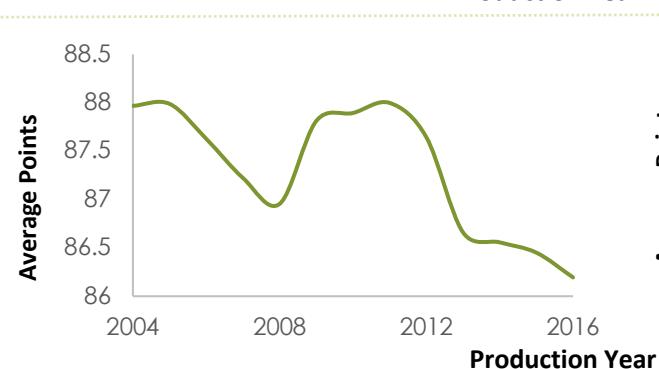
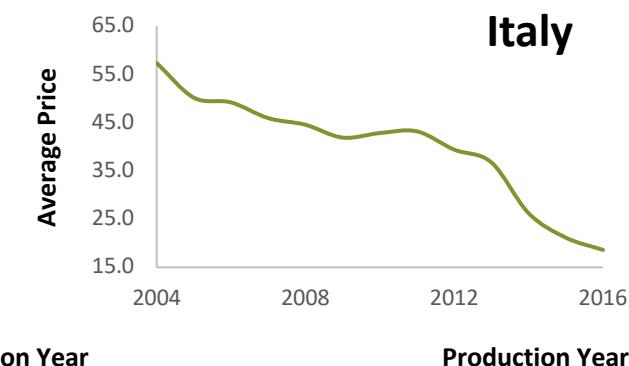
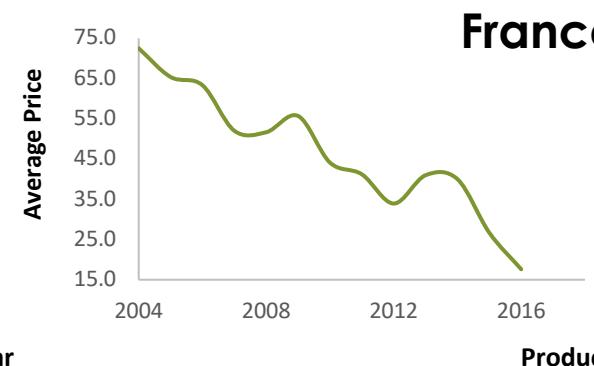
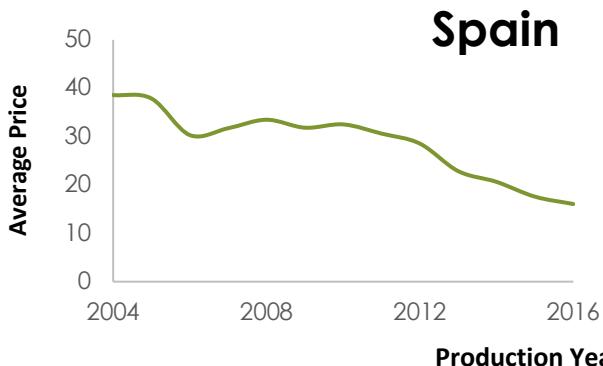
SPARTANS WILL.

Major Business Questions

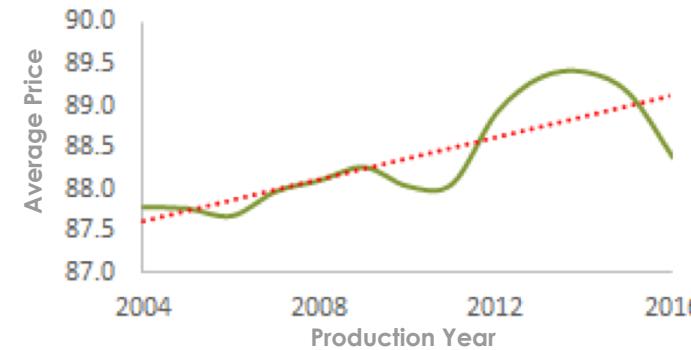
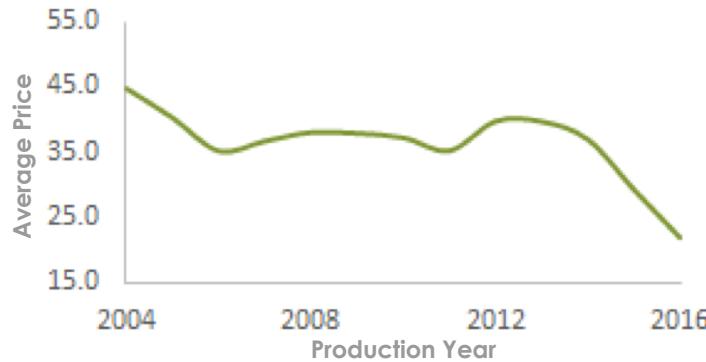
Research Questions

- Are older wines better than the new ones?
- What are the best wines for specific customer segments?
- What is the ideal wine for particular taste preferences?
- What would be the best region to grow the grapes in?

Old is Gold

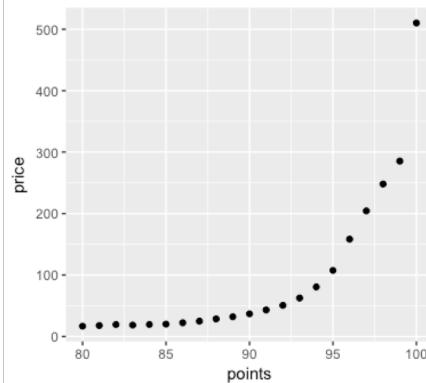


US is the Exception



Wine-making in the US is a very new phenomenon as compared to European countries. Newer wineries are more adept and skilled at making wine as a result

Expensive Wine is Rated Higher – Pricing Model



Coefficients:

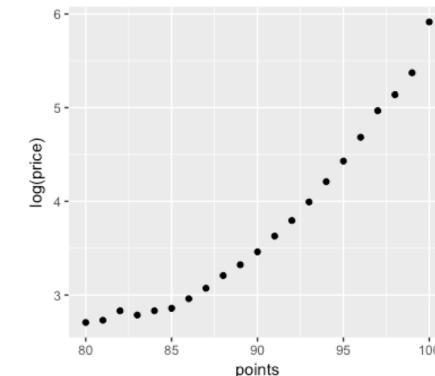
	Estimate	Std. Error	t value	Pr(> t)	Signif. codes:
(Intercept)	-460.46243	3.11657	-147.7	<2e-16 ***	0 '***'
points	5.60750	0.03523	159.2	<2e-16 ***	0.001 '**'
---					0.01 '*'
					0.05 '.'
					0.1 ' '
					1

Residual standard error: 37.3 on 120973 degrees of freedom

(8996 observations deleted due to missingness)

Multiple R-squared: 0.1732, Adjusted R-squared: 0.1732

F-statistic: 2.534e+04 on 1 and 120973 DF, p-value: < 2.2e-16



Observation: Price = f(Points) is exponentially related. “Fine” wines cost significantly more

Pricing Model – Age and Review Score is Significant

$$\text{Price} = f(\text{Points}, \text{Age})$$



Coefficients:

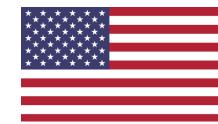
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1600.2682	189.9461	-8.425	1.26e-06 ***
age	1.5421	0.5085	3.033	0.00962 **
pts	18.3730	2.1563	8.521	1.11e-06 ***

Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Residual standard error: 8.278 on 13 degrees of freedom

Multiple R-squared: 0.9128, Adjusted R-squared: 0.8994

F-statistic: 68.03 on 2 and 13 DF, p-value: 1.299e-07



Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-954.550	419.053	-2.278	0.0403 *
age	1.618	0.717	2.256	0.0419 *
pts	11.054	4.686	2.359	0.0346 *

Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Residual standard error: 6.724 on 13 degrees of freedom

Multiple R-squared: 0.3071, Adjusted R-squared: 0.2005

F-statistic: 2.88 on 2 and 13 DF, p-value: 0.09215

Best Wines for Different Customer Segments

Under \$10

Chateau Ste.
Michelle 2011
Riesling - 91

Snoqualmie 2006
Winemaker's Select
Riesling - 91

Herdade dos
Machados 2012
Toutalga Red - 91

\$10-\$25

Rulo 2007 Syrah -
96

Dunham 2010 Lewis
Estate Vineyard
Riesling - 95

Alain Brumont 2010
Château Bouscassé
Red - 95

\$25-\$50

Failla 2010 Estate
Vineyard
Chardonnay - 99

Taylor Fladgate NV
325 Port - 97

Château Coutet
2014 - 97

\$50-\$100

Charles Smith 2006
Royal City Syrah -
100

Cayuse 2008 Bionic
Frog Syrah - 100

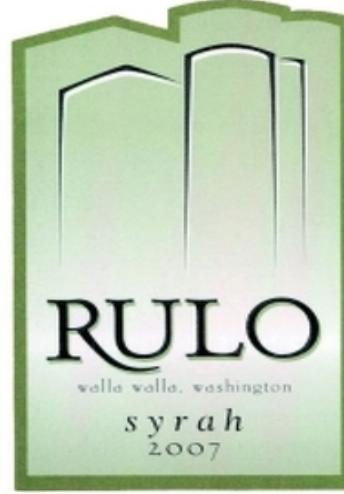
Williams Selyem
2009 Pinot Noir - 99



Best Wines for Different Customer Segments



Under \$10
Rating – 91



\$10-\$25
Rating – 96

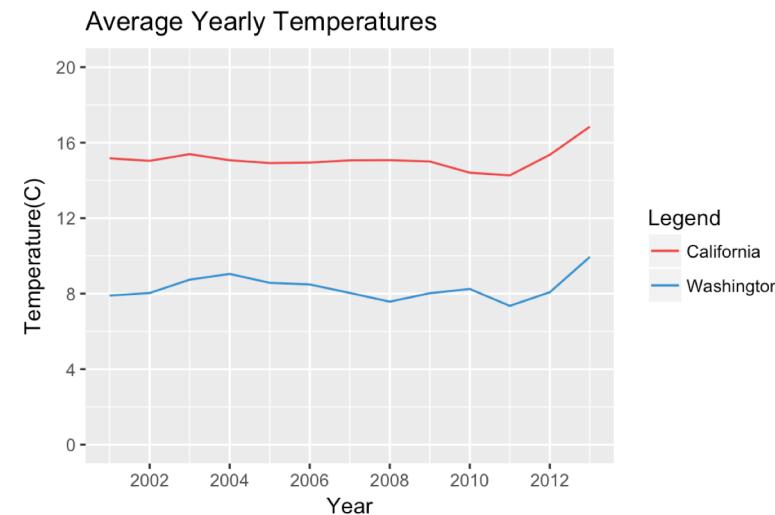
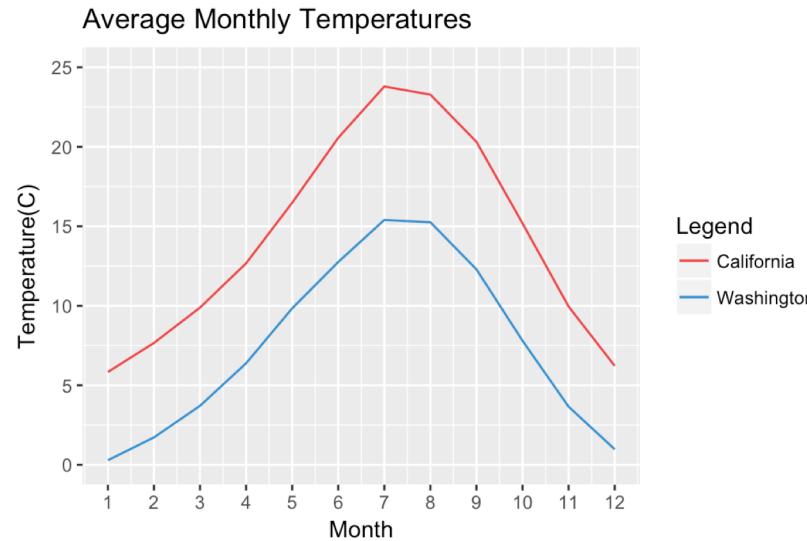


\$25-\$50
Rating - 99



\$50-\$100
Rating - 100

California and Washington Produce Great Wines



- Wine producing grapes required atleast 6 months of 10C-20C temperatures

White Wines – Sweet/Acidic/Fruity

Observation: White wines are great for customers that like fruity, sweet and acidic wines



A word cloud centered around the word "sweet". Other words include "peach", "honey", "apricot", "dessert", "flower", "candied", "soft", "bright", "floral", "creamy", "orange", and "white".

Moscato



A word cloud centered around the word "acidity". Other words include "apple", "pear", "peach", "crisp", "oak", "nose", "ripe", "toast", "citrus", "vanilla", "pineapple", "rich", "lemon", "texture", and "now".

Chardonnay



A word cloud centered around the word "acidity". Other words include "pear", "apple", "peach", "fresh", "spicy", "dry", "sweetness", "lychee", "spice", "notes", "sweet", "rose", "now", "rich", "nose", and "lychee".

Gewurtzaminer



A word cloud centered around the word "pear". Other words include "apple", "acidity", "citrus", "fresh", "light", "ripe", "melon", "spice", "nose", "rich", "dry", "white", "crisp", and "pear".

Pinot Gris

Red Wines – Tart/Oaky/Robust

Observation: Red wines are suitable for customers that like smoky, tart and dry wines

blackberry
chocolate
dark ripe
dry plum
spice acidity nose
fruits oak notes
herbal black
berry cherry

Malbec

black
cherry
blackberry
plum dark now
rich notes soft
cassis currant
oak ripe

Cabernet Sauvignon

now acidity
red cherry
col black
raspberry noir
nose rich spicelight
pinot oak ripe

Pinot Noir

dry plum
chocolate
soft red berry oak
blackberry spice now
rich black notes
cherry

Merlot

Findings

- Older wines are rated higher and cost more in Europe; the trend for American wines is different as the market is newer
- Pricing can be done linearly using the age of the wine and the rating
- White wines are sweeter and fruity on the palate while red wines are darker and tart
- Columbia Valley in Washington is home to some of the best wines in the US due to its even climate



SPARTANS WILL.

Appendix