



IS3100- Techniques for Big Data

Wine Sourcing through Natural Language Processing

Group Penta-thon:

MO Chi Yan (Steven); 55216244
CHOW Ho Keung (Joe); 55474780
FUNG Ka Ho (Cody); 55805390
HO Mei Ting (May); 56222629
LUK Wai Sze (Edith); 56634335





Table of Contents

01

Business Problem

How our project solves the pain point?

02

Data

Explain our prepared dataset.

03

Wine Sourcing

How we apply our prepared dataset in wine sourcing.

04

Other Findings

Other relevant findings discovered in our project.



Contents



1. Business Problem

Scenario: Imagine you are a wine retailer (i.e. Watson's Wine), one of your key business activities is to source wine.

What do you have: List of wines with different features

- Country
- Grape variety
- Winery
- Vintage year
- Price



What is missing: Features that are less standardized in nature, for example, the taste, tannin, texture of wine.

Pain point: As a wine sourcer, you cannot practically know the taste of thousands of millions of wines.

How?





2. Data

Data source: *Kaggle (with ~100,000 reviews scraped from wine review website “Wine Enthusiast”)*

Major data steps:

1. Remove price outliers which are above \$200 USD
2. Extract vintage years from wines
3. Remove rating bias from individual reviewers
4. Extract adjectives describing the wine from reviews





2. Data

Dataset after cleansing and preparation

<i>title</i>	<i>unbiased_points</i>	<i>country</i>	<i>variety</i>	<i>price</i>	<i>description</i>	<i>tokened_description</i>
Quinta dos Avidagos 2011 Avidagos Red (Douro)	85.14	Portugal	Portuguese Red	15	This is ripe and fruity, a wine that is smooth while still structured. Firm tannins are filled out with juicy red berry fruits and freshened with acidity. It's already drinkable, although it will certainly be better from 2016.	['ripe', 'firm', 'red', 'drinkable']
Rainstorm 2013 Pinot Gris (Willamette Valley)	83.99	US	Pinot Gris	14	Tart and snappy, the flavors of lime flesh and rind dominate. Some green pineapple pokes through, with crisp acidity underscoring the flavors. The wine was all stainless-steel fermented.	['snappy', 'flesh', 'green', 'crisp', 'flavor', 'wine']
St. Julian 2013 Reserve Late Harvest Riesling (Lake Michigan Shore)	87	US	Riesling	13	Pineapple rind, lemon pith and orange blossom start off the aromas. The palate is a bit more opulent, with notes of honey-drizzled guava and mango giving way to a slightly astringent, semidry finish.	['blossom', 'opulent', 'honey-drizzled', 'astringent']
Sweet Cheeks 2012 Vintner's Reserve Wild Child Block Pinot Noir (Willamette Valley)	83.99	US	Pinot Noir	65	Much like the regular bottling from 2012, this comes across as rather rough and tannic, with rustic, earthy, herbal characteristics. Nonetheless, if you think of it as a pleasantly unfussy country wine, it's a good companion to a hearty winter stew.	['much', 'regular', 'rough', 'tannic', 'rustic', 'herbal', 'characteristic', 'unfussy', 'good']
Tandem 2011 Ars In Vitro Tempranillo-Merlot (Navarra)	85.29	Spain	Tempranillo-Merlot	15	Blackberry and raspberry aromas show a typical Navarran whiff of green herbs and, in this case, horseradish. In the mouth, this is fairly full bodied, with tomatoey acidity. Spicy, herbal flavors complement dark plum fruit, while the finish is fresh but grabby.	['typical', 'green', 'horseradish', 'full', 'tomatoey', 'herbal', 'finish', 'fresh']





3. Wine Sourcing

Mechanism behind the wine sourcing system



01

Input Segment

Input desired segment based on taste, tannin, texture, and etc.



02

Match Keywords

Compare the adjectives of each wine with keywords in the preset segments, the system will output a list of wines which have sufficient matched keywords.



03

Apply Filters

Based on users' requirements, additional filters like country, grape variety, and price range can be applied to obtain a refined list of wines.



04

Recommend Wine

Recommend top wines in the refined list based on sorting of their unbiased rating points.



3. Wine Sourcing

Example A - Sweet Wine from the US with Melbot grape variety

Choose a wine segment

- Sweet wine
- White wine
- Red wine
- Sparkling wine
- Fruity wine
- Spicy wine
- Mineral Wine

Keywords of Sweet wine (segment)

- Rich
- Sweet
- Luscious
- Dessert-Like
- Honeyed
- Fruity
- Aromatic
- Velvety
- Decadent

Match keywords to each wine review

Keep wines with at least 3 matched keywords with adjectives in the reviews

Further filter wines based on user selections

Country: the US
Variety: Melbot





3. Wine Sourcing

Example A - Sweet Wine from the US with Melbot grape variety

Recommended
Sweet wine



Louis M Martini 1997 Merlot

Result generated from the sourcing machine

		title	points	variety	country	price		description	count_keyword_in
4746	Louis M. Martini	1997 Merlot (Chiles Valley)	88.391973	Merlot	US	25	An exceptional wine from this historic produce...		4
49657	Rodney Strong	1997 Merlot (Sonoma County)	86.391973	Merlot	US	16	Very aromatic for Merlot, with ripe, sugary bl...		3
45689	Belhurst	2007 Traditions Merlot (New York)	85.033810	Merlot	US	27	A winning combination of sweet and savory, Bel...		3

- Sorted by unbiased rating points, number of keywords matched with the selected segment





3. Wine Sourcing

Example B - Spicy Wine from France with price < \$30 USD

Choose a wine
segment

- Spicy wine
- Sweet wine
- White wine
- Red wine
- Sparkling wine
- Fruity wine
- Mineral Wine

Keywords of Sweet wine
(segment)

- Peppery
- Spicy
- Bold
- Intense
- Warm
- Rich
- Full-bodied
- Exotic
- Aromatic

Match keywords to each wine review

Keep wines with at least 3 matched
keywords with adjectives in the reviews

Further filter wines based on user selections

Country: France

Price: Below \$30 USD





3. Wine Sourcing

Example B- Spicy Wine from France with price < \$30 USD

Recommended
Spicy wine



Château d'Aqueria 2014 Tavel

Result generated from the sourcing machine

	title	points	variety	country	price	description	count_keyword_in
61960	Château d'Aqueria 2014 Tavel	89.394490	Rosé	France	20	Full-bodied, rich and slightly warm, this asse...	4
7417	Pierre Sparr 2006 Reserve Gewurztraminer (Alsace)	84.391973	Gewürztraminer	France	17	Those who seek out Gewürztraminer for its inte...	4

- Sorted by unbiased rating points, number of keywords matched with the selected segment





4. Other Findings



Countries' Key Features

country	0	1	2	3	4	5	6	7	8	9
Germany	lemon-lime	delicate	lime	off-dry	lemon	brisk	floral	zesty	luscious	sunny
Austria	freshness	pear	lemon	aromatic	easy					
Italy	underbrush	wild	smooth	fine-grained	sangiovese	cherry	vibrant	tannin	exotic	
Australia	silky	pear	complex	crisp	medium-bodied	next	supple	olive		
South Africa	approachable	unfold	medium-weight	assertive	easy	lush				
New Zealand	silky	citrusy	medium-bodied	next	noir	supple	new			
France	young	tight	generous	give	ready					
Portugal	young	powerful	smooth	generous	touriga	nacional	ready	give	tannin	
US	complex	currant	new							
Spain	basic	toasty	tight	raw						
Argentina	raw	toasty	floral							
Chile	basic	olive								

German Wine

- Acidic
- Light texture



Australian Wine

- Sour
- Heavier texture
- Fruitful



4. Other Findings



Grape Varieties' Key Features

variety	0	1	2	3	4	5	6	7	8
Nebbiolo	underbrush	firm	raspberry	herb	sour	fine-grained	assertive	balsamic	spice
<i>Riesling</i>	lemon-lime	lime	off-dry	residual	brisk	floral	long		
<i>Grüner Veltliner</i>	yeasty	concentrated	intense	savory					
<i>Rhône-style Red Blend</i>	silky	medium-bodied	rhône-style	dusty	olive				
<i>Pinot Noir</i>	noir	silky							
<i>Portuguese Red</i>	powerful	old	touriga	nacional	impressive	give	dense		
<i>Syrah</i>	olive								
<i>Sangiovese</i>	underbrush	fleshy	fine-grained	blue	assertive	black-skinned	balsamic	sangiovese	spice
<i>Bordeaux-style Red Blend</i>	powerful	potential	bordeaux-style	black-currant	give				
<i>Cabernet Sauvignon</i>									
<i>Red Blend</i>	sangiovese								
<i>Chardonnay</i>	toast	pineapple							
<i>Cabernet Franc</i>	bramble								
<i>Malbec</i>	deep	toasty	oaky	heavy					
<i>Pinot Gris</i>	off-dry	residual	bitter	gris	pinot				
<i>Zinfandel</i>	classic	hot	high	briary	rustic	old			
<i>Tempranillo</i>	raw	toasty	tight	oaky	heavy				
<i>White Blend</i>	apricot	floral							
Sauvignon Blanc	grassy	medium-bodied	lime						
<i>Merlot</i>	nice								
<i>Rosé</i>	rosé	tight	intense	pale	strawberry				

Nebbiolo

- Herb flavor
- Spice

Sauvignon Blanc

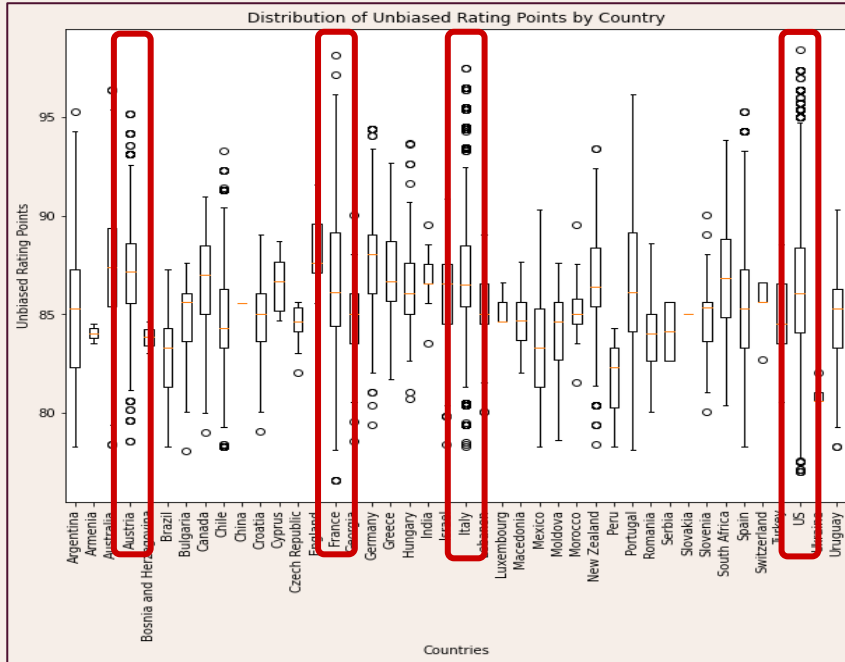
- Sour
- Heavier texture
- Fruitful



4. Other Findings

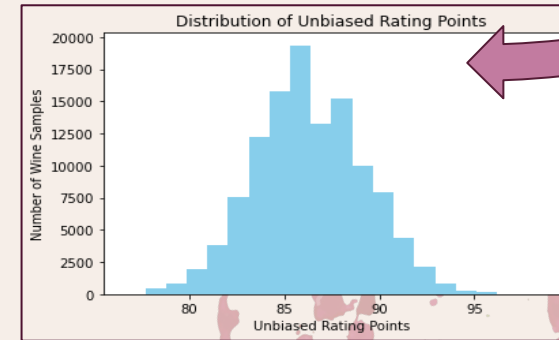


Countries' Wine Unbiased Rating Points Distribution



- A. Austria, France, Italy, US, etc.
- > Wide spectrum of wines
 - > More divergence in wine quality

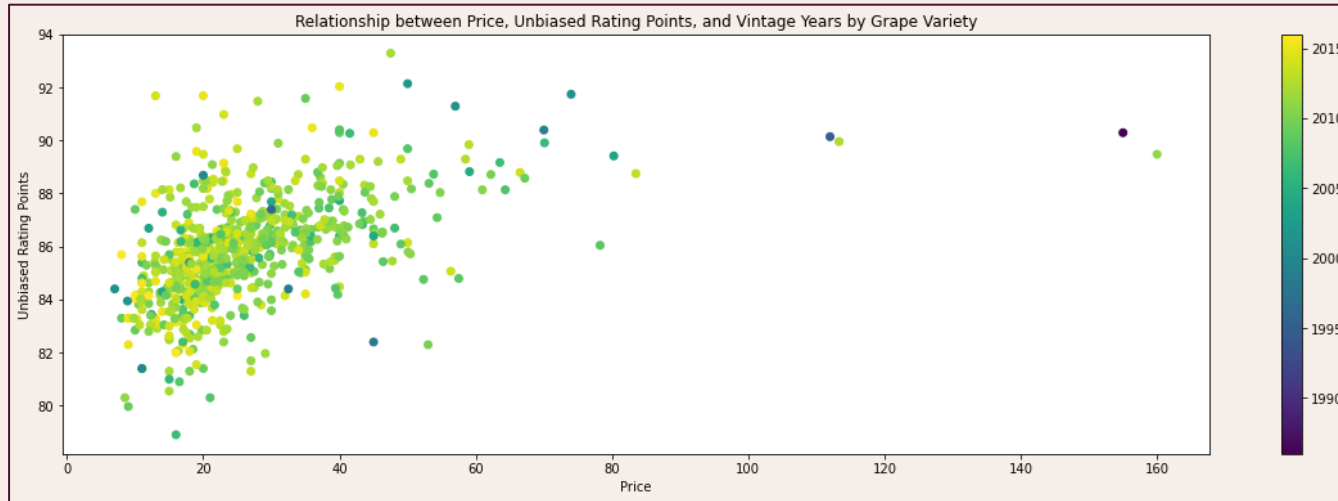
B. Overall unbiased rating points follow a rather normal distribution



4. Other Findings



Relationship between Price, Unbiased Rating Points, and Vintage Years by Grape Variety



- A. Positive relationship between price and unbiased rating points according to each grape variety
- B. No clear relationship between vintage years, price, and unbiased rating points
-> Not as an important factor in determining the price and quality of a particular grape variety

Thanks!

Q&A Session

