Digital Market Analysis for Red Wine in US

Foo Suon Chuang (Bryan), Miao Li

Overview

- Problem Statement
  - What characteristics are associated with good wines?
  - Application: designing successful product, framing attractive marketing messages
- Consumers
  - Professional sommelier
  - Average Wine Drinkers
- Two Web-scraped Datasets
  - Data scraped from *Wine Enthusiast*, one of the most authoritative and largest wine review websites
  - Data scraped from *Vivino*, the world's largest wine marketplace with reviews straight from the community
- Measurement of Wine Quality
  - Wine Spectator's 100-Point Scale
    - **95-100 Classic:** a great wine
    - **90-94 Outstanding:** a wine of superior character and style
    - **85-89 Very good:** a wine with special qualities
    - **80-84 Good:** a solid, well-made wine
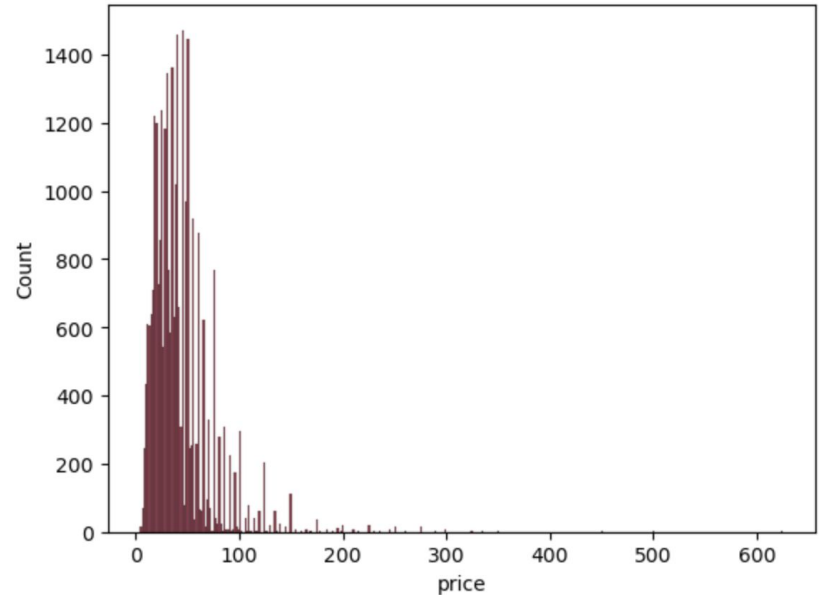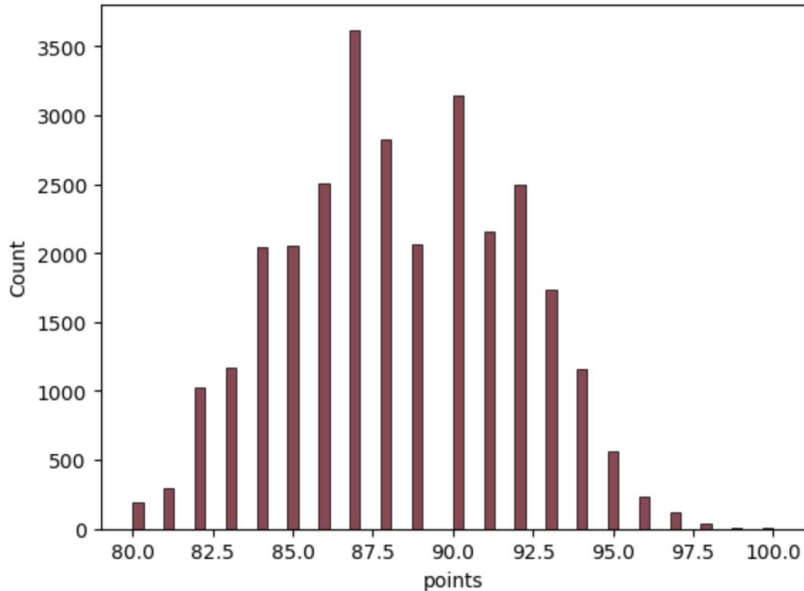 - Community Reviews

Industry Landscape

Data exploration to understand red wines in US and zooming-in on California

Dataset 1: *Wine Enthusiast* Reviews

- Contains points and reviews from professional wine reviewers for 29,468 red wines produced in the US

Cleaned Wine Reviews to get keywords in description

- Tokenize reviews by removing special characters and stopwords and lemmatizing, then converting tokens to 2-grams

	description	des_token	2_gram
0	This tremendous 100% varietal wine hails from Oakville and was aged over three years in oak. Juicy red-cherry fruit and a compelling hint of caramel greet the palate, framed by elegant, fine tannins and a subtle minty tone in the background. Balanced and rewarding from start to finish, it has ye...	[tremendous, varietal, wine, hail, oakville, aged, three, year, oak, juicy, fruit, compelling, hint, caramel, greet, palate, framed, elegant, fine, tannin, subtle, minty, tone, background, balanced, rewarding, start, finish, year, ahead, develop, nuance, enjoy]	[(tremendous, varietal), (varietal, wine), (wine, hail), (hail, oakville), (oakville, aged), (aged, three), (three, year), (year, oak), (oak, juicy), (juicy, fruit), (fruit, compelling), (compelling, hint), (hint, caramel), (caramel, greet), (greet, palate), (palate, framed), (framed, elegant), ...
3	This spent 20 months in 30% new French oak, and incorporates fruit from Ponzi's Aurora, Abetina and Madrona vineyards, among others. Aromatic, dense and toasty, it deftly blends aromas and flavors of toast, cigar box, blackberry, black cherry, coffee and graphite. Tannins are polished to a fine ...	[spent, month, new, french, oak, incorporates, fruit, ponzi, aurora, abetina, madrona, vineyard, among, others, aromatic, dense, toasty, deftly, blend, aroma, flavor, toast, cigar, box, blackberry, black, cherry, coffee, graphite, tannin, polished, fine, sheen, frame, finish, loaded, dark, choco...	[(spent, month), (month, new), (new, french), (french, oak), (oak, incorporates), (incorporates, fruit), (fruit, ponzi), (ponzi, aurora), (aurora, abetina), (abetina, madrona), (madrona, vineyard), (vineyard, among), (among, others), (others, aromatic), (aromatic, dense), (dense, toasty), (toast...
8	This re-named vineyard was formerly bottled as deLancellotti. You'll find striking minerality underscoring chunky black fruits. Accents of citrus and graphite comingle, with exceptional midpalate concentration. This is a wine to cellar, though it is already quite enjoyable. Drink now through 2030.	[vineyard, formerly, bottled, delancellotti, find, striking, minerality, underscoring, chunky, black, fruit, accent, citrus, graphite, comingle, exceptional, midpalate, concentration, wine, cellar, though, already, quite, enjoyable, drink]	[(vineyard, formerly), (formerly, bottled), (bottled, delancellotti), (delancellotti, find), (find, striking), (striking, minerality), (minerality, underscoring), (underscoring, chunky), (chunky, black), (black, fruit), (fruit, accent), (accent, citrus), (citrus, graphite), (graphite, comingle),...
9	The producer sources from two blocks of the vineyard for this wine—one at a high elevation, which contributes bright acidity. Crunchy cranberry, pomegranate and orange peel flavors surround silky, succulent layers of texture that present as fleshy fruit. That delicately lush flavor has considera...	[producer, source, two, block, vineyard, high, elevation, contributes, bright, acidity, crunchy, cranberry, pomegranate, orange, peel, flavor, surround, silky, succulent, layer, texture, present, fleshy, fruit, delicately, lush, flavor, considerable, length]	[(producer, source), (source, two), (two, block), (block, vineyard), (vineyard, high), (high, elevation), (elevation, contributes), (contributes, bright), (bright, acidity), (acidity, crunchy), (crunchy, cranberry), (cranberry, pomegranate), (pomegranate, orange), (orange, peel), (peel, flavor),...
11	From 18-year-old vines, this supple well-balanced effort blends flavors of mocha, cherry, vanilla and breakfast tea. Superbly integrated and delicious even at this early stage, this wine seems destined for a long and savory cellar life. Drink now through 2028.	[vine, supple, effort, blend, flavor, mocha, cherry, vanilla, breakfast, tea, superbly, integrated, delicious, even, early, stage, wine, seems, destined, long, savory, cellar, life, drink]	[(vine, supple), (supple, effort), (effort, blend), (blend, flavor), (flavor, mocha), (mocha, cherry), (cherry, vanilla), (vanilla, breakfast), (breakfast, tea), (tea, superbly), (superbly, integrated), (integrated, delicious), (delicious, even), (even, early), (early, stage), (stage, wine), (wi...

What is most discussed when reviewing red wine
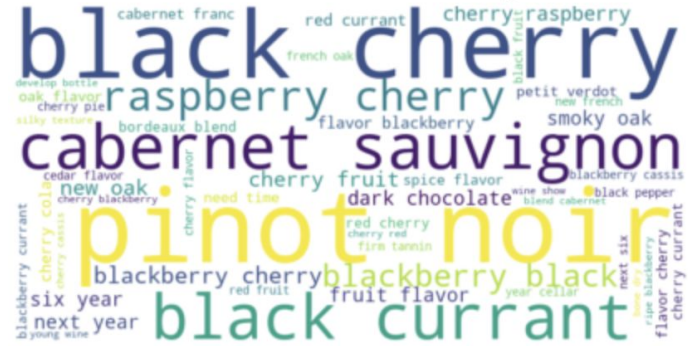
Dataset 2: *Vivino* User Reviews

- **145 red wines** from 91 wineries in California, USA

- Data on average rating, number of reviews, price, characteristics, flavours, varietal, region

	wineryName	name	year	acidity	intensity	sweetnes	tannin	price	ratingAverage	numberTasterReviews	totalRatings	region	varietal
0	Paul Hobbs	Paul Hobbs Beckstoffer Las Piedras Vineyard Ca...	2015	3.371048	4.626113	1.805029	3.323693	346.95	4.8	68	86	Alexander Valley	Cabernet Sauvignon
1	Bernardus	Bernardus Marinus 2013	2013	3.743659	4.761420	1.572299	3.692956	34.99	4.3	51	98	Alexander Valley	Cabernet Sauvignon
2	Beau Vigne	Beau Vigne Legacy Cabernet Sauvignon 2020	2020	3.343543	4.570415	1.643999	3.484988	59.99	4.2	88	69	Napa Valley	Cabernet Sauvignon
3	DAOU	DAOU Reserve Cabernet Sauvignon 2021	2021	3.059963	4.795351	1.668042	3.384822	39.99	4.3	510	689	Paso Robles	Cabernet Sauvignon
4	Paul Hobbs	Paul Hobbs Beckstoffer Las Piedras Vineyard Ca...	2015	3.371048	4.626113	1.805029	3.323693	346.95	4.8	68	86	Alexander Valley	Cabernet Sauvignon
...	...	...	...	...	...	...	...	...	...	...	...	...	...
147	Beau Vigne	Beau Vigne Romeo Cabernet Sauvignon 2021	2021	3.355670	4.587563	1.728072	3.344321	99.99	4.4	133	1227	Napa Valley	Cabernet Sauvignon
148	J Dusi	J Dusi Paper Street The Narrator 2021	2021	4.038571	4.721000	1.538543	3.542857	29.99	4.4	20	158	California	Blend
149	Château St. Jean	Château St. Jean Cinq Cépages 2017	2017	3.532186	4.797547	1.654068	3.604700	69.99	4.4	140	327	California	Blend
150	Sbragia	Sbragia Neal Vineyard Cabernet Sauvignon 2017	2017	3.492442	4.980658	1.131783	3.660969	59.99	4.4	15	119	Napa Valley	Cabernet Sauvignon
151	Beringer	Beringer Private Reserve Cabernet Sauvignon 2017	2017	3.355557	4.690163	1.705016	3.432122	129.99	4.4	853	345	Napa Valley	Cabernet Sauvignon

152 rows × 14 columns

Dataset 2: *Vivino* User Reviews

- 4 main red wine groups:

  - "Cabernet Sauvignon", "Pinot Noir", "Zinfandel", "Blend" (encompassing all different types of red blends such as Rhone blends, Bordeaux blends, ect.)

- 8 different wine regions in California:

  - "California" (wine made of grapes coming from all parts of the state), "Alexander Valley", "Napa valley", "Russian River Valley", " Paso Robles", "Santa Barbara", "Sonoma County", "Sonoma Coast"

varietal	varietal
Blend	31
Cabernet Sauvignon	96
Pinot Noir	13
Zinfandel	6

region	region
Alexander Valley	13
California	39
Napa Valley	74
Paso Robles	7
Russian River Valley	2
Santa Barbara	3
Sonoma Coast	3
Sonoma County	5

8

Brand Perception

- Selected the most-rated winery in California, taking the average of characteristics of wines produced there
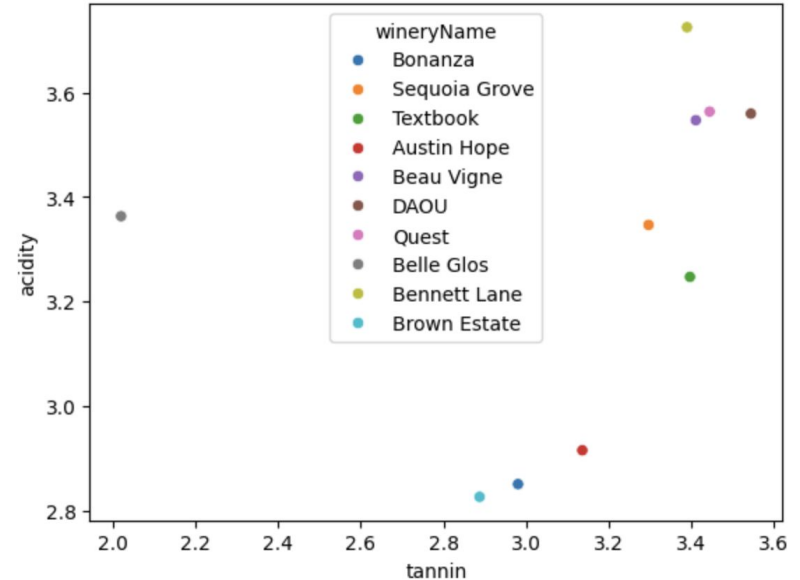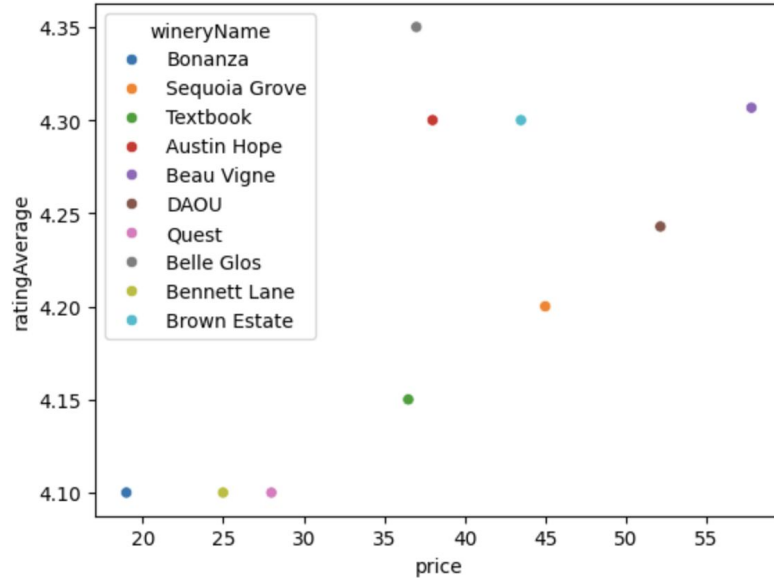
Professional Wine Reviews Sentiment Analysis

What distinguishes 'Classic' wines from 'Good' wines for professionals

Sentiment Analysis of *Wine Enthusiast* Reviews

- Data:
  - Randomly sampled Good wines (0) to combine with Classic wines (1) in Wine Enthusiast dataset to have a balanced dataset

- Analysis
  - Construct TF-IDF features matrix from cleaned wine reviews
  - Conduct logistic regression to find coefficients for n-grams

Sentiment Analysis of *Wine Enthusiast* Reviews

"Good" Wine (80-84 points)

cherry flavor	-1.659013
blackberry cherry	-0.924580
berry flavor	-0.914928
soft simple	-0.817385
high alcohol	-0.732436
blackberry cherry flavor	-0.717096
cola flavor	-0.673062
taste like	-0.671813
cherry cola flavor	-0.626051
petite sirah	-0.615358

"Classic" Wine (95-100 points)

black currant	1.581592
dark chocolate	1.207377
next year	1.206974
cabernet sauvignon	1.176565
many year	0.974072
blackberry black	0.960374
new oak	0.938573
asian spice	0.883611
blackberry black currant	0.793690
six year	0.7834

Consumer Preferences

How do average consumers on *Vivino* like different wine characteristics

Regression for *Vivino* User Reviews

- **Outcome variable: Average User Rating Score (out of 5)**

  - Mean rating: 4.28

  - Minimum rating: 3.7

  - Maximum rating: 4.8

- **18 regressor variables:**

  - 5 wine characteristics

  - 13 wine flavour categories

Regressors: 5 Wine Characteristics

● "Acidity", "Intensity", "Sweetness", "Tannin", "Price"

	wineryName	name	year	acidity	intensity	sweetnes	tannin	price
0	Paul Hobbs	Paul Hobbs Beckstoffer Las Piedras Vineyard Ca...	2015	3.371048	4.626113	1.805029	3.323693	346.95
1	Bernardus	Bernardus Marinus 2013	2013	3.743659	4.761420	1.572299	3.692956	34.99
2	Beau Vigne	Beau Vigne Legacy Cabernet Sauvignon 2020	2020	3.343543	4.570415	1.643999	3.484988	59.99
3	DAOU	DAOU Reserve Cabernet Sauvignon 2021	2021	3.059963	4.795351	1.668042	3.384822	39.99
4	Paul Hobbs	Paul Hobbs Beckstoffer Las Piedras Vineyard Ca...	2015	3.371048	4.626113	1.805029	3.323693	346.95
...	...	...	...	...	...	...	...	...
147	Beau Vigne	Beau Vigne Romeo Cabernet Sauvignon 2021	2021	3.355670	4.587563	1.728072	3.344321	99.99
148	J Dusi	J Dusi Paper Street The Narrator 2021	2021	4.038571	4.721000	1.538543	3.542857	29.99
149	Château St. Jean	Château St. Jean Cinq Cépages 2017	2017	3.532186	4.797547	1.654068	3.604700	69.99
150	Sbragia	Sbragia Neal Vineyard Cabernet Sauvignon 2017	2017	3.492442	4.980658	1.131783	3.660969	59.99
151	Beringer	Beringer Private Reserve Cabernet Sauvignon 2017	2017	3.355557	4.690163	1.705016	3.432122	129.99

Regressors: 13 Wine Flavour Categories

- Total of 13 flavour categories extracted across the top 9 flavour categories for with the largest number of mentions for each wine: "oak", 'black_fruit', "earth", "red_fruit", "spices", "non_oak", "microbio", "floral", "tree_fruit", "citrus_fruit", "dried_fruit", "tropical_fruit", "vegetal"

- Calculate the **proportion of reviews in each wine** that had mentioned a given flavour category

	uid	oak	black_fruit	earth	red_fruit	spices	non_oak	microbio	floral	tree_fruit	citrus_fruit	dried_fruit	tropical_fruit	vegetal
0	a59994c3-33a4-4605-8bf9-20a5b9a2723b	0.476744	0.453488	0.441860	0.244186	0.302326	0.348837	0.058140	0.034884	0.0	0.000000	0.034884	0.000000	0.000000
1	57a7b820-e103-41d3-8a69-9050d7f63a94	1.071429	0.775510	0.867347	0.734694	0.561224	0.683673	0.142857	0.000000	0.0	0.000000	0.112245	0.000000	0.081633
2	38e8893b-3abb-4af8-981d-913d11e43d47	0.869565	0.724638	0.202899	0.536232	0.333333	0.347826	0.086957	0.028986	0.0	0.000000	0.057971	0.000000	0.000000
3	f97093e4-e922-4b1a-9485-caeeb140b2c5	0.915820	0.686502	0.551524	0.447025	0.422351	0.519594	0.075472	0.029028	0.0	0.000000	0.076923	0.000000	0.000000
4	df269b9c-085f-4d0b-a62f-4e347274d02a	0.476744	0.453488	0.441860	0.244186	0.302326	0.348837	0.058140	0.034884	0.0	0.000000	0.034884	0.000000	0.000000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
141	44a8886c-b799-4095-a3f5-24a44fdc0cea	0.095355	0.074980	0.061125	0.039120	0.043195	0.061125	0.011410	0.003260	0.0	0.002445	0.000000	0.000000	0.000000
142	8661b4d1-7e37-4184-847e-a8ff779f22e0	0.069620	0.075949	0.031646	0.075949	0.025316	0.025316	0.000000	0.006329	0.0	0.000000	0.012658	0.006329	0.000000
143	3c12d123-f757-4309-9fdf-1f8cd81ba32e	1.440367	1.226300	1.033639	0.749235	0.568807	0.877676	0.204893	0.064220	0.0	0.000000	0.094801	0.000000	0.000000
144	39a37c36-a80f-44f7-acc3-6fd1f12ff8cb	0.025210	0.016807	0.016807	0.016807	0.008403	0.016807	0.000000	0.000000	0.0	0.000000	0.008403	0.000000	0.000000
145	00152d36-15dd-486a-9847-859f0a029ac6	2.600000	2.171014	1.802899	0.823188	1.153623	1.550725	0.228986	0.115942	0.0	0.000000	0.173913	0.000000	0.000000

146 rows × 14 columns

Fixed Effects: Year and Region

- Two sets of fixed effects:

  - Year: Dummy variable columns to control for differences in year in which the wine was produced

 - Region: Dummy variable columns to control for differences on where the wine was produced

uid	2013	2015	2016	2017	2018	2019	2020	2021	2022
a59994c3-33a4-4605-8bf9-20a5b9a2723b	0	1	0	0	0	0	0	0	0
57a7b820-e103-41d3-8a69-9050d7f63a94	1	0	0	0	0	0	0	0	0
38e8893b-3abb-4af8-981d-913d11e43d47	0	0	0	0	0	0	1	0	0
f97093e4-e922-4b1a-9485-caeeb140b2c5	0	0	0	0	0	0	0	1	0
df269b9c-085f-4d0b-a62f-4e347274d02a	0	1	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...
44a8886c-b799-4095-a3f5-24a44fdc0cea	0	0	0	0	0	0	0	1	0
8661b4d1-7e37-4184-847e-a8ff779f22e0	0	0	0	0	0	0	0	1	0
3c12d123-f757-4309-9fdf-1f8cd81ba32e	0	0	0	1	0	0	0	0	0
39a37c36-a80f-44f7-acc3-6fd1f12ff8cb	0	0	0	1	0	0	0	0	0
00152d36-15dd-486a-9847-859f0a029ac6	0	0	0	1	0	0	0	0	0

uid	Alexander Valley	California	Napa Valley	Paso Robles	Russian River Valley	Santa Barbara	Sonoma Coast	Sonoma County
a59994c3-33a4-4605-8bf9-20a5b9a2723b	1	0	0	0	0	0	0	0
57a7b820-e103-41d3-8a69-9050d7f63a94	1	0	0	0	0	0	0	0
38e8893b-3abb-4af8-981d-913d11e43d47	0	0	1	0	0	0	0	0
f97093e4-e922-4b1a-9485-caeeb140b2c5	0	0	0	1	0	0	0	0
df269b9c-085f-4d0b-a62f-4e347274d02a	1	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...
44a8886c-b799-4095-a3f5-24a44fdc0cea	0	0	1	0	0	0	0	0
8661b4d1-7e37-4184-847e-a8ff779f22e0	0	1	0	0	0	0	0	0
3c12d123-f757-4309-9fdf-1f8cd81ba32e	0	1	0	0	0	0	0	0
39a37c36-a80f-44f7-acc3-6fd1f12ff8cb	0	0	1	0	0	0	0	0
00152d36-15dd-486a-9847-859f0a029ac6	0	0	1	0	0	0	0	0

145 rows × 8 columns

Regression Results for *Vivino* User Reviews

	Statistically Significant Variables	Coefficients
Wine Characteristics	Acidity	**0.029** **
	Intensity	**0.045** **
	Tannin	**0.0326tannin^2 - 0.266tannin** *
	Price	**0.001** ***
Wine Flavour Categories	Red Fruit	**-0.029** .
	Dried Fruit	**0.114** *
Region	California	**-0.036** *

Input

$0.0326\,x^2 - 0.266\,x$

Plots

(x from −0.8 to 9)

(x from −44.9 to 53)
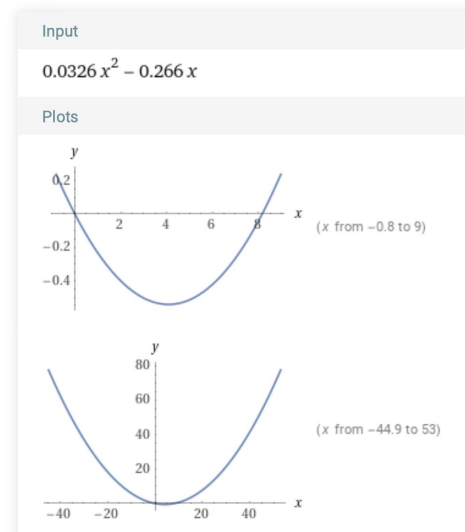
Aggregated Conjoint Analysis on *Vivino* User Reviews

- Data Processing

  - Aggregated total-market data from Vivino, getting ratings on a scale of 5

  - Creating our own profile and level data with 4 attributes

  - Selected attributes with the most samples

- 123 wines with 16 Levels:

  - **Region**: Napa Valley, California, Alexander Valley

  - **Grape Varietal**: Blend, Cabernet Sauvignon, Pinot Noir, Zinfandel

  - **Prominent flavor**: blackfruit, earth, oak, red fruit

  - **Secondary Flavor**: blackfruit, earth, oak, red fruit, spices

Aggregated Conjoint Analysis on *Vivino* User Reviews

	name	region	varietal	flavor_1	flavor_2
1	Paul Hobbs Beckstoffer Las Piedras Vineyard Cabernet...	Alexander Valley	Cabernet Sauvignon	oak	black_fruit
2	Bernardus Marinus 2013	Alexander Valley	Cabernet Sauvignon	oak	earth
3	Beau Vigne Legacy Cabernet Sauvignon 2020	Napa Valley	Cabernet Sauvignon	oak	black_fruit
4	Paul Hobbs Beckstoffer Las Piedras Vineyard Cabernet...	Alexander Valley	Cabernet Sauvignon	oak	black_fruit
5	Bonanza Cabernet Sauvignon Lot	California	Cabernet Sauvignon	oak	black_fruit
6	Bennett Lane Maximus 2018	California	Blend	oak	black_fruit
7	Rombauer Vineyards Cabernet Sauvignon 2019	Napa Valley	Cabernet Sauvignon	oak	black_fruit
8	Nuxy Impulse Cabernet Sauvignon 2020	California	Cabernet Sauvignon	red_fruit	oak

Aggregated Conjoint Analysis on *Vivino* User Reviews

Region		
	Alexander Valley	0.119
	California	-0.153
	Napa Valley	0.034
Grape Varietal	Blend	0.022
	Cabernet Sauvignon	-0.031
	Pinot Noir	0.034
	Zinfandel	-0.025

Prominent Flavor		
	Black fruit	0.084
	Earth	-0.011
	Oak	0.031
	Red fruit	-0.104
Secondary Flavor	Black fruit	0.048
	Earth	--0.029
	Oak	0.023
	Red fruit	-0.068
	Spices	0.026

Insights and Recommendations

- Product Design and Positioning

    - Wines **produced in specific wine regions** are more popular than wines that are produced from just "California" (grapes that come from all parts of the state).

    - Wines with **some ageing potential** (based on the "next year" token in *Wine Enthusiast* reviews) tend to receive higher scores from professional wine reviewers.

    - Wines with **either low or high amounts of tannin** are more popular among the average consumer. When tannin is at a moderate level (neither low nor high), they tend to correlate with lower user ratings.

Insights and Recommendations

- Wine Characteristics

    - **Black fruit (eg. black currant) and oak (dark chocolate) flavours** in wines tend to receive higher scores by both professional wine reviewers and average consumers.

    - **Red fruit and earth flavours** are less popular among average consumers, but more popular among professional wine reviewers.

Limitations

- Vivino User Review Scores Data

  - **Lack of individual-level consumer profile data** to perform more detailed market segmentation and part worth utilities analysis (only available after paywall).

  - **Lack of time-series sales data** for each wine to better understanding changes in sales vis-a-vis competitor wines (only available after paywall).

  - Web-scraped **data size is relatively small for some varietals and regions**. It would have been better if we had a larger sample sizes of some of the red wine varietals (such as for "Pinot Noir" and "Zinfandel") and some of the wine regions (such as "Russian River Valley" and "Santa Barbara").

- Wine Enthusiast Reviews Data

  - Most wine ratings by Wine Enthusiast are at least 'good wine' – **lack of wines that obtain scores below 80 points**.

Thank you!