

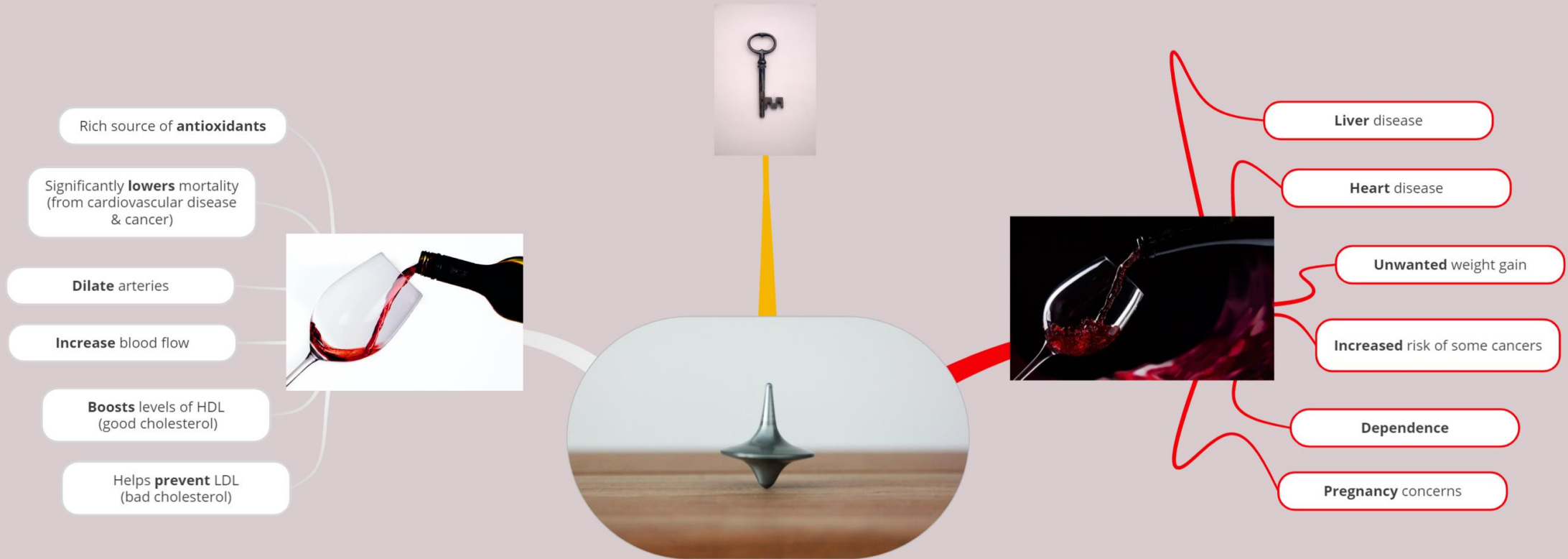
A low-angle, upward-looking photograph of a diverse group of graduates in black caps and gowns. They are gathered outdoors, with lush green trees in the background. The graduates are smiling and looking upwards, some with their arms raised in celebration. The text "Summative Capstone" is overlaid in white, centered on the image.

Summative Capstone

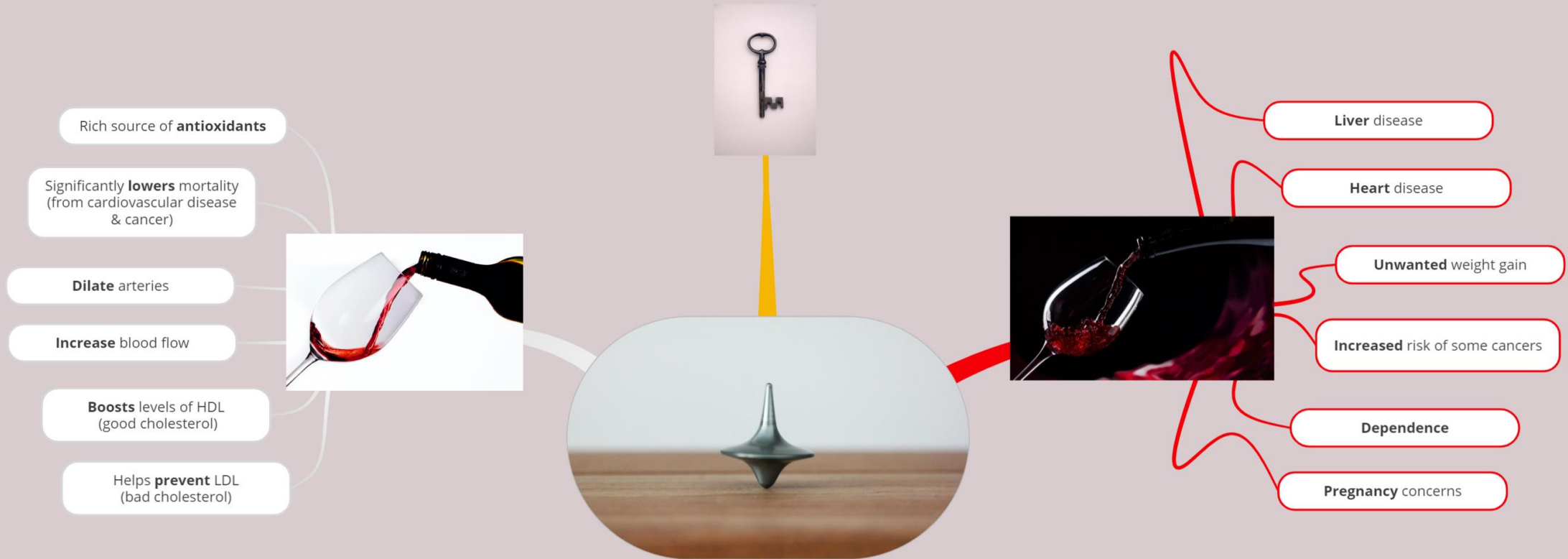


Traditional (western) medicine

Benefits & Risks - The key to a balancing act...



Benefits & Risks - The key to a balancing act...



1 standard glass (150ml) a day

How to select ?

- Chardonnay
- Sauvignon Blanc
- Pinot Blanc
- Pinot Gris / Pinot Grigio
- Semillon
- Gewurztraminer
- Riesling
- Viognier
- Chenin Blanc
- Gruener Veltliner



- Cabernet Sauvignon
- Merlot
- Zinfandel
- Syrah / Shiraz
- Malbec
- Pinot Noir
- Sangiovese
- Nebbiolo
- Bordeaux
- Rhone



- Champagne
- Prosecco
- Moscato



- Blended
- Saignée
- Designate

1



1



2



1



2



3



1



2



3



4

100

1



2



3



4

100

5



1



2



3



4



5



Column Name		Data Type
Number		int
country		nvarchar(50)
description	3	nvarchar(MAX)
designation		nvarchar(MAX)
points	4	int
price	5	float
province		nvarchar(MAX)
region_1		nvarchar(MAX)
region_2		nvarchar(MAX)
taster_name	1	nvarchar(50)
taster_twitter_handle		nvarchar(50)
title	2	nvarchar(MAX)
variety		nvarchar(MAX)
winery		nvarchar(MAX)

1



2



3



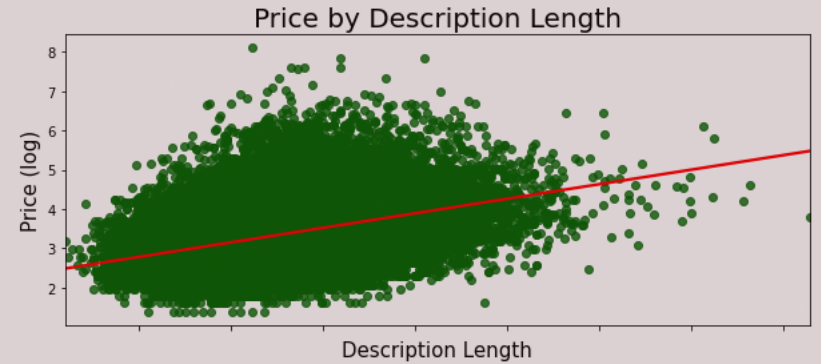
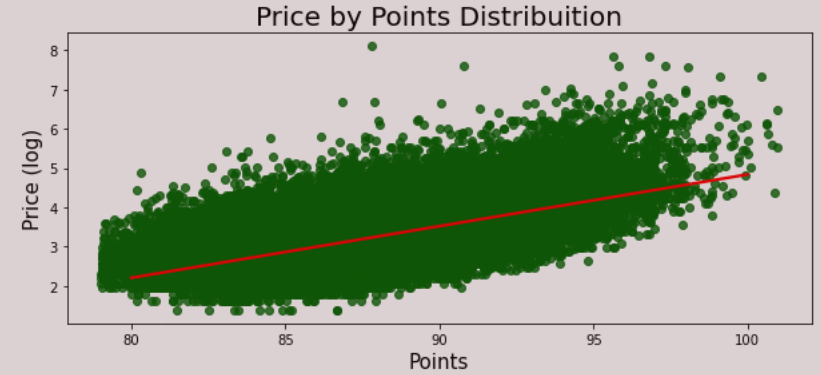
4

100

5

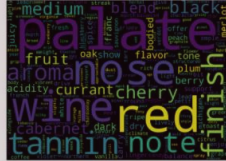


Column Name		Data Type
Number		int
country		nvarchar(50)
description	3	nvarchar(MAX)
designation		nvarchar(MAX)
points	4	int
price	5	float
province		nvarchar(MAX)
region_1		nvarchar(MAX)
region_2		nvarchar(MAX)
taster_name	1	nvarchar(50)
taster_twitter_handle		nvarchar(50)
title	2	nvarchar(MAX)
variety		nvarchar(MAX)
winery		nvarchar(MAX)



We do not need the tasters individually...

Alexander Peartree



Anna Lee C. Iijima



Anne Krebiehl



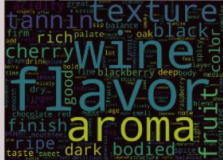
Carrie Dykes



Fiona Adams



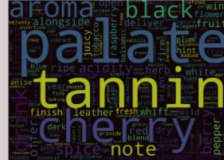
Jim Gordon



Joe Czerwinski



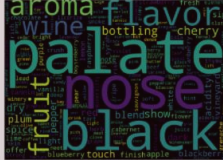
Kerin OKeefe



Lauren Buzzeo



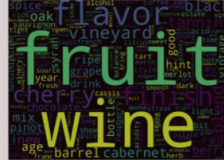
Matt Kettmann



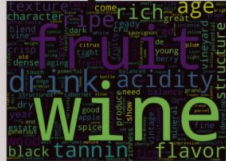
Michael Schachner



Paul Gregutt



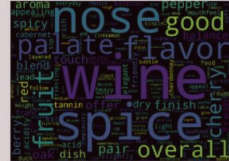
Roger Voss



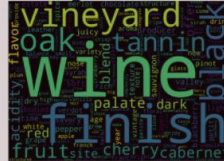
Sean P. Sullivan



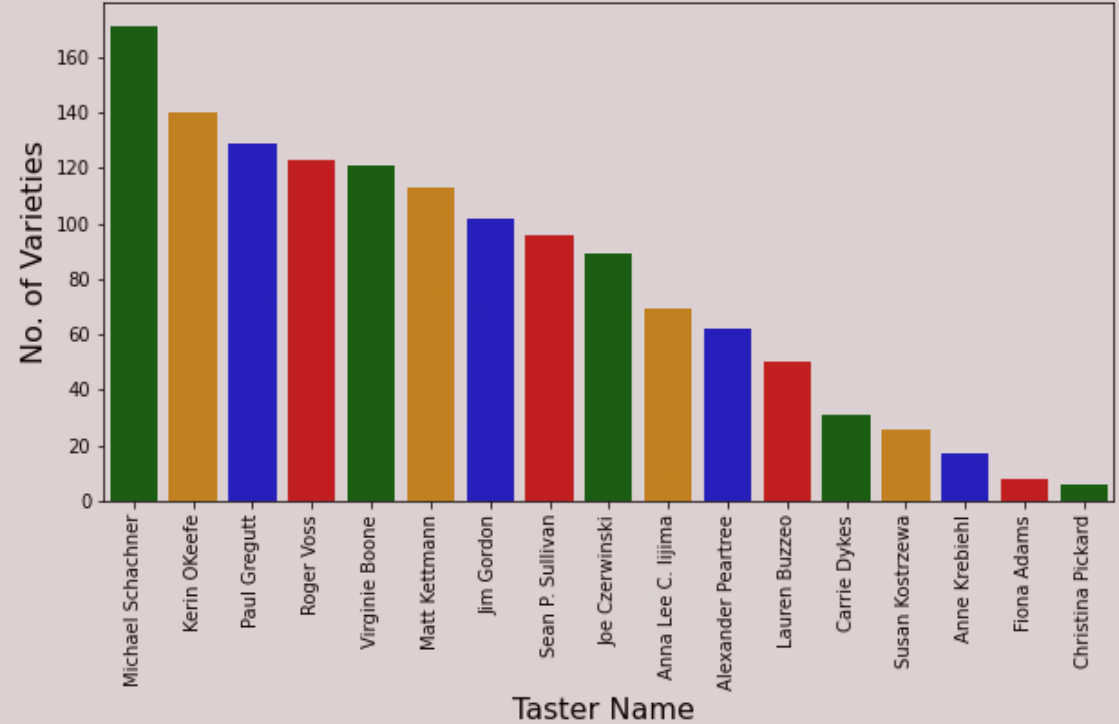
Susan Kostzewa



Virginie Boone



Number of Varieties Rated By Different Tasters



We need them as a group !

We do not need the tasters individually...

Alexander Peartree



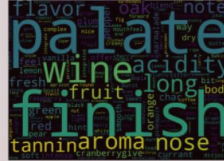
Anna Lee C. Iijima



Anne Krebiehl



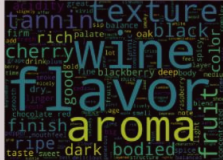
Carrie Dykes



Fiona Adams



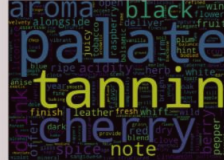
Jim Gordon



Joe Czerwinski



Kerin OKeefe



Lauren Buzzeo



Matt Kettmann



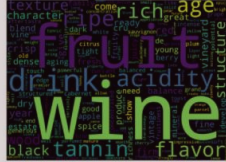
Michael Schachner



Paul Gregutt



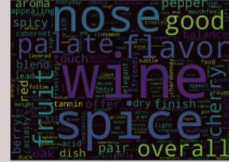
Roger Voss



Sean P. Sullivan



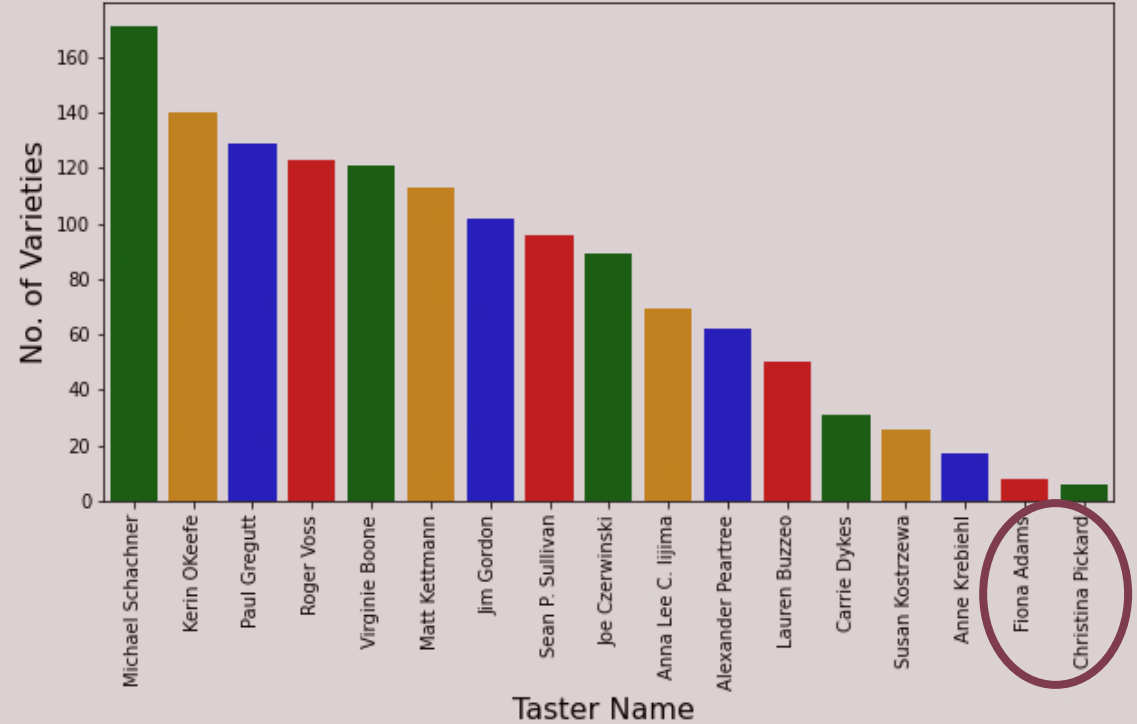
Susan Kostzewa



Virginie Boone



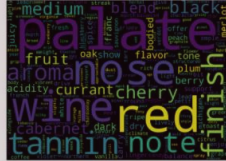
Number of Varieties Rated By Different Tasters



We need them as a group !

We do not need the tasters individually...

Alexander Peartree



Anna Lee C. Iijima



Anne Krebiehl



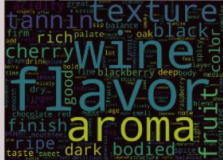
Carrie Dykes



Fiona Adams



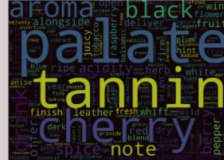
Jim Gordon



Joe Czerwinski



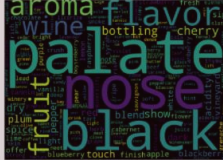
Kerin OKeefe



Lauren Buzzeo



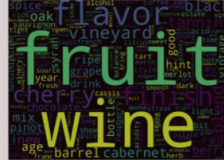
Matt Kettmann



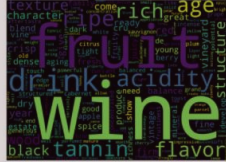
Michael Schachner



Paul Gregutt



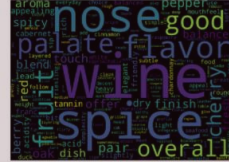
Roger Voss



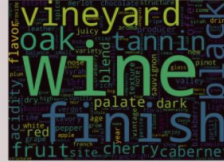
Sean P. Sullivan



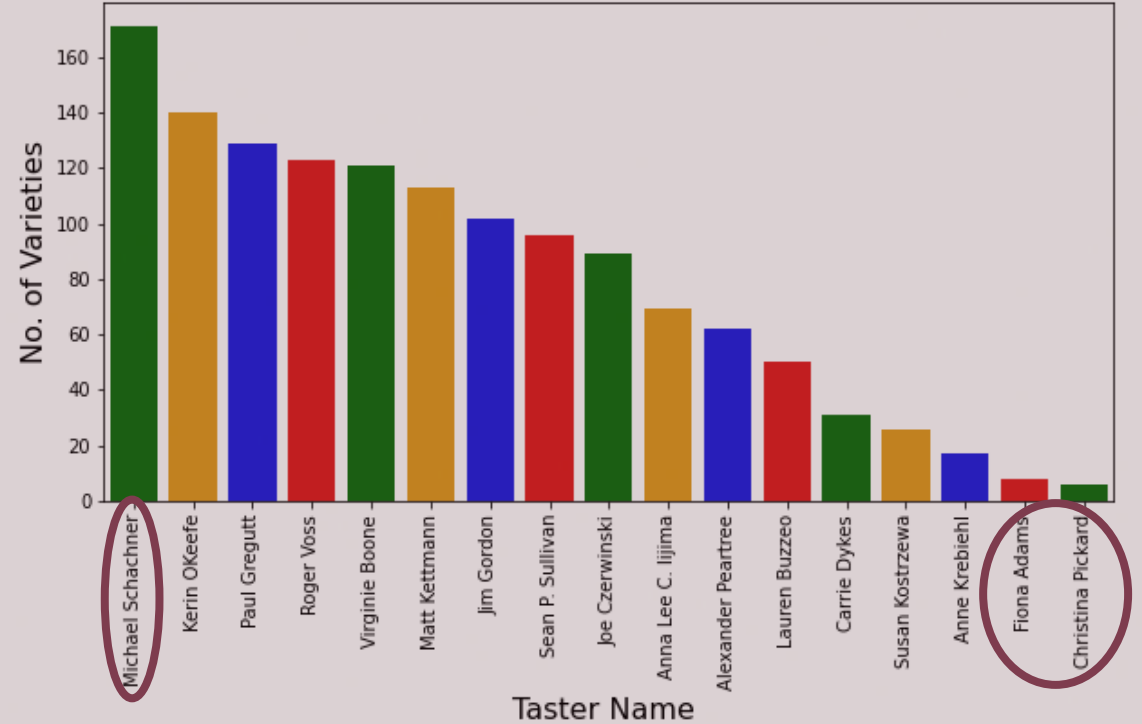
Susan Kostzewa



Virginie Boone



Number of Varieties Rated By Different Tasters



We need them as a group !

Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)

Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)

Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)



Most common *descriptive* word (top 20)

Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)



Most common *descriptive* word (top 20):

*fruit, cherry, spice, oak, berry,
plum, apple, blackberry, citrus, vanilla,
pepper, lemon, raspberry, peach, chocolate,
pear, fresh, rich, dry, sweet.*

Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)



Most common *descriptive* word (top 20):

*fruit, cherry, spice, oak, berry,
plum, apple, blackberry, citrus, vanilla,
pepper, lemon, raspberry, peach, chocolate,
pear, fresh, rich, dry, sweet.*



Results

	ModelName	ModelNumber	StartTime	EndTime	Duration	UnitOfTime	Accuracy Score
1	Decision Tree	Gini Depth:3	2021-08-14 22:38:36.600	2021-08-14 22:38:37.593	0.991151	Seconds	0.1332
2	Decision Tree	Gini Depth:5	2021-08-14 22:38:41.733	2021-08-14 22:38:41.917	0.183183	Seconds	0.1482
3	Decision Tree	Gini Depth:10	2021-08-14 22:38:41.980	2021-08-14 22:38:42.200	0.220325	Seconds	0.1712
4	Decision Tree	Gini Depth:25	2021-08-14 22:38:42.260	2021-08-14 22:38:42.543	0.281806	Seconds	0.179



For any emergency, please call 9 wine wine



Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)



Most common *descriptive* word (top 20):

*fruit, cherry, spice, oak, berry,
plum, apple, blackberry, citrus, vanilla,
pepper, lemon, raspberry, peach, chocolate,
pear, fresh, rich, dry, sweet.*



Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)



Most common *descriptive* word (top 20):

*fruit, cherry, spice, oak, berry,
plum, apple, blackberry, citrus, vanilla,
pepper, lemon, raspberry, peach, chocolate,
pear, fresh, rich, dry, sweet.*



Most common (top 20):

Down from >700

```
('Pinot Noir', 12278),  
( 'Chardonnay', 10868),  
( 'Cabernet Sauvignon', 8840),  
( 'Red Blend', 8243),  
( 'Bordeaux-style Red Blend', 6471),  
( 'Riesling', 4773),  
( 'Sauvignon Blanc', 4575),  
( 'Syrah', 3828),  
( 'Rosé', 3220),  
( 'Merlot', 2896),  
( 'Nebbiolo', 2607),  
( 'Zinfandel', 2530),  
( 'Sangiovese', 2462),  
( 'Malbec', 2446),  
( 'Portuguese Red', 2282),  
( 'White Blend', 2179),  
( 'Sparkling Blend', 1971),  
( 'Tempranillo', 1671),  
( 'Rhône-style Red Blend', 1343),  
( 'Pinot Gris', 1324)]
```

Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)



Sentence cleaning (Lemmatization):



Most common (top 20):

Down from >700

```
('Pinot Noir', 12278),
('Chardonnay', 10868),
('Cabernet Sauvignon', 8840),
('Red Blend', 8243),
('Bordeaux-style Red Blend', 6471),
('Riesling', 4773),
('Sauvignon Blanc', 4575),
('Syrah', 3828),
('Rosé', 3220),
('Merlot', 2896),
('Nebbiolo', 2607),
('Zinfandel', 2530),
('Sangiovese', 2462),
('Malbec', 2446),
('Portuguese Red', 2282),
('White Blend', 2179),
('Sparkling Blend', 1971),
('Tempranillo', 1671),
('Rhône-style Red Blend', 1343),
('Pinot Gris', 1324)]
```

```
# Sentence cleaning:
def preproc_cleantext(text_list):
    documents = []
    stemmer = WordNetLemmatizer()
    spell = Speller(lang='en')
    for sen in range(0, len(text_list)):
        # Remove all the special characters
        document = re.sub(r'\W', ' ', str(text_list[sen]))
        # remove all single characters
        document = re.sub(r'\s+[a-zA-Z]\s+', ' ', document)
        # Remove single characters from the start
        document = re.sub(r'^\s+[a-zA-Z]\s+', ' ', document)
        # Substituting multiple spaces with single space
        document = re.sub(r'\s+', ' ', document, flags=re.I)
        # Removing prefixed 'b'
        document = re.sub(r'^b\s+', '', document)
        # Converting to Lowercase
        document = document.lower()
        # regular expression: 50 kg (this step return a list of word in each sentence)
        document = re.findall(r'[0-9]+|[a-z]+', document)

        ## get a list of word in each sentence:
        # document = document.split()
        # Spelling Correction
        document = [spell(word) for word in document]
        # print(document)
        # Lemmatization: reduce the word into dictionary root form
        document = [stemmer.lemmatize(word) for word in document]
        document = ' '.join(document)
        documents.append(document)
    return documents
```

Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)



Sentence cleaning (Lemmatization):

```
description_list = preproc_cleantext(description_list)
```

```
start: 20210803 09:24:28  
--end: 20210803 12:46:58  
timing in seconds: 12150.633126  
timing in minutes: 202.51
```



Most common (top 20):

Down from >700

```
('Pinot Noir', 12278),  
( 'Chardonnay', 10868),  
( 'Cabernet Sauvignon', 8840),  
( 'Red Blend', 8243),  
( 'Bordeaux-style Red Blend', 6471),  
( 'Riesling', 4773),  
( 'Sauvignon Blanc', 4575),  
( 'Syrah', 3828),  
( 'Rosé', 3220),  
( 'Merlot', 2896),  
( 'Nebbiolo', 2607),  
( 'Zinfandel', 2530),  
( 'Sangiovese', 2462),  
( 'Malbec', 2446),  
( 'Portuguese Red', 2282),  
( 'White Blend', 2179),  
( 'Sparkling Blend', 1971),  
( 'Tempranillo', 1671),  
( 'Rhône-style Red Blend', 1343),  
( 'Pinot Gris', 1324)]
```

```
# Sentence cleaning:  
  
def preproc_cleantext(text_list):  
    documents = []  
    stemmer = WordNetLemmatizer()  
    spell = Speller(lang='en')  
    for sen in range(0, len(text_list)):  
        # Remove all the special characters  
        document = re.sub(r'\W', ' ', str(text_list[sen]))  
        # remove all single characters  
        document = re.sub(r'\s+[a-zA-Z]\s+', ' ', document)  
        # Remove single characters from the start  
        document = re.sub(r'\^[a-zA-Z]\s+', ' ', document)  
        # Substituting multiple spaces with single space  
        document = re.sub(r'\s+', ' ', document, flags=re.I)  
        # Removing prefixed 'b'  
        document = re.sub(r'^b\s+', '', document)  
        # Converting to Lowercase  
        document = document.lower()  
        # regular expression: 50 kg (this step return a list of word in each sentence)  
        document = re.findall(r'[0-9]+|[a-z]+', document)  
  
        ## get a list of word in each sentence:  
        # document = document.split()  
        # Spelling Correction  
        document = [spell(word) for word in document]  
        # print(document)  
        # Lemmatization: reduce the word into dictionary root form  
        document = [stemmer.lemmatize(word) for word in document]  
        document = ' '.join(document)  
        documents.append(document)  
    return documents
```

Results

	ModelName	ModelNumber	StartTime	EndTime	Duration	UnitOfTime	AccuracyScore
1	Decision Tree	Gini Depth:3	2021-08-14 22:38:36.600	2021-08-14 22:38:37.593	0.991151	Seconds	0.1332
2	Decision Tree	Gini Depth:5	2021-08-14 22:38:41.733	2021-08-14 22:38:41.917	0.183183	Seconds	0.1482
3	Decision Tree	Gini Depth:10	2021-08-14 22:38:41.980	2021-08-14 22:38:42.200	0.220325	Seconds	0.1712
4	Decision Tree	Gini Depth:25	2021-08-14 22:38:42.260	2021-08-14 22:38:42.543	0.281806	Seconds	0.179
5	Linear SVM	SGD Classifier 5	2021-08-15 07:17:24.307	2021-08-15 08:57:40.833	100.28	Minutes	0.637

Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)



Sentence cleaning (Lemmatization):

```
description_list = preproc_cleantext(description_list)
```

```
start: 20210803 09:24:28  
--end: 20210803 12:46:58  
timing in seconds: 12150.633126  
timing in minutes: 202.51
```



Most common (top 20):

Down from >700

```
('Pinot Noir', 12278),  
( 'Chardonnay', 10868),  
( 'Cabernet Sauvignon', 8840),  
( 'Red Blend', 8243),  
( 'Bordeaux-style Red Blend', 6471),  
( 'Riesling', 4773),  
( 'Sauvignon Blanc', 4575),  
( 'Syrah', 3828),  
( 'Rosé', 3220),  
( 'Merlot', 2896),  
( 'Nebbiolo', 2607),  
( 'Zinfandel', 2530),  
( 'Sangiovese', 2462),  
( 'Malbec', 2446),  
( 'Portuguese Red', 2282),  
( 'White Blend', 2179),  
( 'Sparkling Blend', 1971),  
( 'Tempranillo', 1671),  
( 'Rhône-style Red Blend', 1343),  
( 'Pinot Gris', 1324)]
```

```
# Sentence cleaning:  
  
def preproc_cleantext(text_list):  
    documents = []  
    stemmer = WordNetLemmatizer()  
    spell = Speller(lang='en')  
    for sen in range(0, len(text_list)):  
        # Remove all the special characters  
        document = re.sub(r'\W', ' ', str(text_list[sen]))  
        # remove all single characters  
        document = re.sub(r'^[a-zA-Z]\s+', ' ', document)  
        # Remove single characters from the start  
        document = re.sub(r'^[a-zA-Z]\s+', ' ', document)  
        # Substituting multiple spaces with single space  
        document = re.sub(r'\s+', ' ', document, flags=re.I)  
        # Removing prefixed 'b'  
        document = re.sub(r'^b\s+', ' ', document)  
        # Converting to Lowercase  
        document = document.lower()  
        # regular expression: 50 kg (this step return a list of word in each sentence)  
        document = re.findall(r'[0-9]+|[a-z]+', document)  
  
        ## get a list of word in each sentence:  
        # document = document.split()  
        # Spelling Correction  
        document = [spell(word) for word in document]  
        # print(document)  
        # Lemmatization: reduce the word into dictionary root form  
        document = [stemmer.lemmatize(word) for word in document]  
        document = ' '.join(document)  
        documents.append(document)  
    return documents
```

Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)



Sentence cleaning (Lemmatization) +
Latent Dirichlet Allocation (LDA)



Most common (top 20):

Down from >700

```
('Pinot Noir', 12278),  
( 'Chardonnay', 10868),  
( 'Cabernet Sauvignon', 8840),  
( 'Red Blend', 8243),  
( 'Bordeaux-style Red Blend', 6471),  
( 'Riesling', 4773),  
( 'Sauvignon Blanc', 4575),  
( 'Syrah', 3828),  
( 'Rosé', 3220),  
( 'Merlot', 2896),  
( 'Nebbiolo', 2607),  
( 'Zinfandel', 2530),  
( 'Sangiovese', 2462),  
( 'Malbec', 2446),  
( 'Portuguese Red', 2282),  
( 'White Blend', 2179),  
( 'Sparkling Blend', 1971),  
( 'Tempranillo', 1671),  
( 'Rhône-style Red Blend', 1343),  
( 'Pinot Gris', 1324)]
```

Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)



Sentence cleaning (Lemmatization) +
Latent Dirichlet Allocation (LDA)



	text	intent
123119	The wine is seriously structured for the vinta...	Bordeaux-style Red Blend
62381	The wood is too much for the fruit with this w...	Chardonnay
87310	One of Chile's benchmark wines is typically cl...	Cabernet Sauvignon
57262	Simple and soft, with buttery citrus flavors. ...	Sauvignon Blanc
93387	Seeming delicate at first, barely-there whiffs...	Riesling



Most common (top 20):

Down from >700

```
('Pinot Noir', 12278),  
( 'Chardonnay', 10868),  
( 'Cabernet Sauvignon', 8840),  
( 'Red Blend', 8243),  
( 'Bordeaux-style Red Blend', 6471),  
( 'Riesling', 4773),  
( 'Sauvignon Blanc', 4575),  
( 'Syrah', 3828),  
( 'Rosé', 3220),  
( 'Merlot', 2896),  
( 'Nebbiolo', 2607),  
( 'Zinfandel', 2530),  
( 'Sangiovese', 2462),  
( 'Malbec', 2446),  
( 'Portuguese Red', 2282),  
( 'White Blend', 2179),  
( 'Sparkling Blend', 1971),  
( 'Tempranillo', 1671),  
( 'Rhône-style Red Blend', 1343),  
( 'Pinot Gris', 1324)]
```

Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)



Sentence cleaning (Lemmatization) +
Latent Dirichlet Allocation (LDA)



	text	intent
123119	The wine is seriously structured for the vinta...	Bordeaux-style Red Blend
62381	The wood is too much for the fruit with this w...	Chardonnay
87310	One of Chile's benchmark wines is typically cl...	Cabernet Sauvignon
57262	Simple and soft, with buttery citrus flavors. ...	Sauvignon Blanc
93387	Seeming delicate at first, barely-there whiffs...	Riesling



Most common (top 8)

Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)



Sentence cleaning (Lemmatization) +
Latent Dirichlet Allocation (LDA)



	text	intent
123119	The wine is seriously structured for the vinta...	Bordeaux-style Red Blend
62381	The wood is too much for the fruit with this w...	Chardonnay
87310	One of Chile's benchmark wines is typically cl...	Cabernet Sauvignon
57262	Simple and soft, with buttery citrus flavors. ...	Sauvignon Blanc
93387	Seeming delicate at first, barely-there whiffs...	Riesling



Most common (top 8) +
uniform training dataset

Back to the drawing board...

Column Name	Data Type
Number	int
country	nvarchar(50)
description	nvarchar(MAX)
designation	nvarchar(MAX)
points	int
price	float
province	nvarchar(MAX)
region_1	nvarchar(MAX)
region_2	nvarchar(MAX)
taster_name	nvarchar(50)
taster_twitter_handle	nvarchar(50)
title	nvarchar(MAX)
variety	nvarchar(MAX)
winery	nvarchar(MAX)



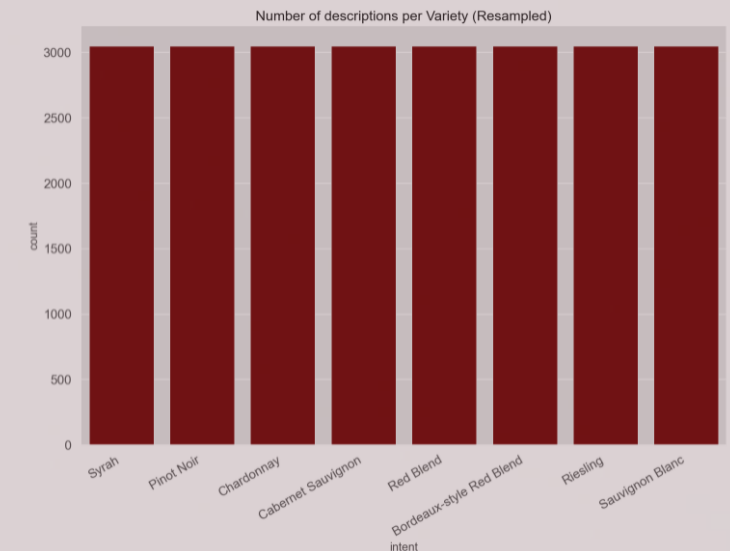
Sentence cleaning (Lemmatization) +
Latent Dirichlet Allocation (LDA)



	text	intent
123119	The wine is seriously structured for the vinta...	Bordeaux-style Red Blend
62381	The wood is too much for the fruit with this w...	Chardonnay
87310	One of Chile's benchmark wines is typically cl...	Cabernet Sauvignon
57262	Simple and soft, with buttery citrus flavors. ...	Sauvignon Blanc
93387	Seeming delicate at first, barely-there whiffs...	Riesling

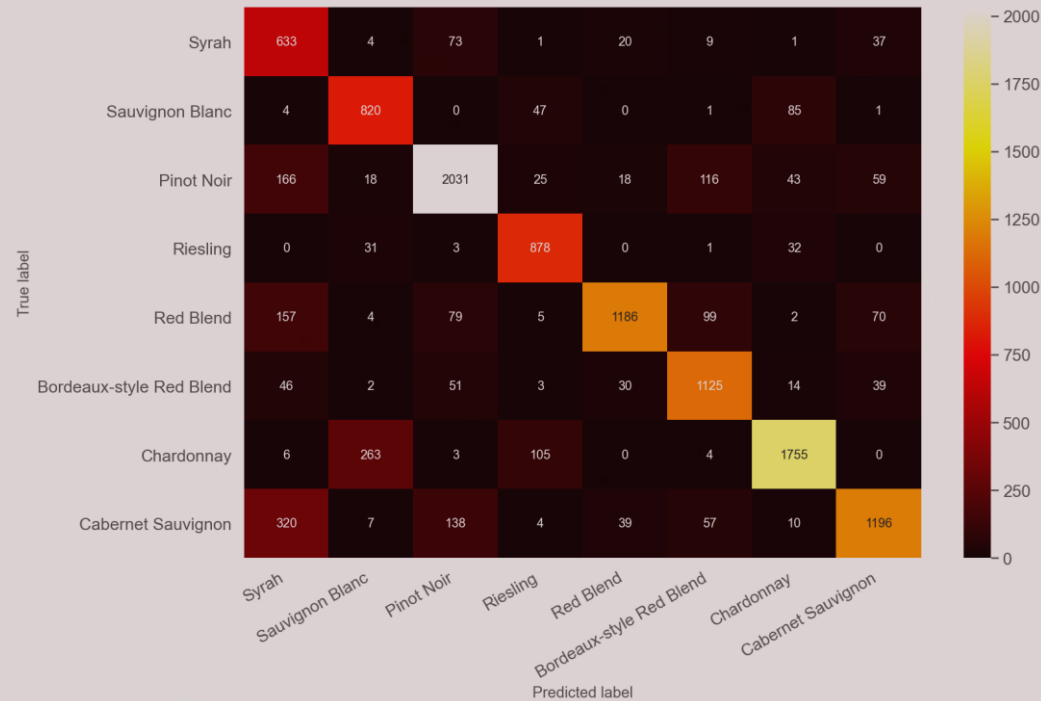


Most common (top 8) +
uniform training dataset



Results

	ModelName	ModelNumber	StartTime	EndTime	Duration	UnitOfTime	AccuracyScore
1	Decision Tree	Gini Depth:3	2021-08-14 22:38:36.600	2021-08-14 22:38:37.593	0.991151	Seconds	0.1332
2	Decision Tree	Gini Depth:5	2021-08-14 22:38:41.733	2021-08-14 22:38:41.917	0.183183	Seconds	0.1482
3	Decision Tree	Gini Depth:10	2021-08-14 22:38:41.980	2021-08-14 22:38:42.200	0.220325	Seconds	0.1712
4	Decision Tree	Gini Depth:25	2021-08-14 22:38:42.260	2021-08-14 22:38:42.543	0.281806	Seconds	0.179
5	Linear SVM	SGD Classifier 5	2021-08-15 07:17:24.307	2021-08-15 08:57:40.833	100.28	Minutes	0.637
6	Linear SVM	LDA 1	2021-08-14 23:22:13.167	2021-08-15 15:19:10.050	956.95	Minutes	0.8036



What we were looking for !



text: heavy wine made of white grapes
intent: Sauvignon Blanc

text: heavy wine made of red grapes
intent: Red Blend

text: strong fruit taste
intent: Bordeaux-style Red Blend

text: vanilla oak
intent: Cabernet Sauvignon

text: lemon fresh
intent: Riesling

text: citrus rich
intent: Sauvignon Blanc

text: apple
intent: Chardonnay

text: pepper
intent: Syrah

text: cherry raspberry dry light
intent: Pinot Noir

text: peach pear
intent: Riesling

Future opportunities...

K-Nearest Neighbors

Recommendation for ## Cabernet Franc ##:

- 1: Malbec with distance: 0.5113234472149908
- 2: Bordeaux-style Red Blend with distance: 0.524946925867908
- 3: Cabernet Sauvignon with distance: 0.5584257196676218
- 4: Cabernet Blend with distance: 0.5696962684504496
- 5: Viognier with distance: 0.5907609970039982

Recommendation for ## Viognier-Chardonnay ##:

- 1: Negrette with distance: 0.47638792563384613
- 2: Tannat-Syrah with distance: 0.47863677781036773
- 3: Merlot-Malbec with distance: 0.48227109511202526
- 4: Syrah-Cabernet with distance: 0.4989966538907661
- 5: Gros and Petit Manseng with distance: 0.5507630663236116

Recommendation for ## Vermentino ##:

- 1: Provence red blend with distance: 0.4586088241743389
- 2: Pinot-Chardonnay with distance: 0.4646322584407232
- 3: Roussanne-Viognier with distance: 0.4646918710937282
- 4: Arneis with distance: 0.46486706039175774
- 5: Alicante with distance: 0.46772470571869995

Recommendation for ## Cabernet Sauvignon ##:

- 1: Merlot with distance: 0.4482167687888756
- 2: Bordeaux-style Red Blend with distance: 0.46709871923047863
- 3: Syrah with distance: 0.4910379903641744
- 4: Red Blend with distance: 0.5405605153575976
- 5: Cabernet Franc with distance: 0.5584257196676218

Recommendation for ## Lambrusco Salamino ##:

- 1: Trebbiano Spoletino with distance: 0.0
- 2: Lambrusco Grasparossa with distance: 0.0
- 3: Lambrusco di Sorbara with distance: 0.0
- 4: Cesanese d'Affile with distance: 0.0
- 5: Malvasia di Candia with distance: 0.0



Cosine similarity

a wine you liked: Skinner 2013 Petite Sirah (El Dorado)

*****Here are a few recommendations for you*****

- 1 . Six Sigma Ranch 2012 Else's Reserve Cabernet Sauvignon (Lake County)
- 2 . Chacewater 2013 Made with Organically Grown Grapes Syrah (Sierra Foothills)
- 3 . Concannon 2014 Captain Joe's Petite Sirah (Livermore Valley)
- 4 . King Estate 1999 Croft Vineyard Pinot Noir (Willamette Valley)
- 5 . Ferrari-Carano 2013 Pinot Noir (Anderson Valley)
- 6 . Pirouette 2013 Vintage Select Solo Dancer Red (Columbia Valley (WA))
- 7 . The Farm Winery 2013 LPF Cabernet Sauvignon (Adelaida District)
- 8 . Sextant 2013 El Pomar Estate Cabernet Sauvignon (El Pomar District)
- 9 . William Hatcher 2014 Pinot Noir (Willamette Valley)
- 10 . Grassini 2012 Cabernet Sauvignon (Happy Canyon of Santa Barbara)



Q & A

Sources

<https://www.kaggle.com/zynicide/wine-reviews?select=winemag-data-130k-v2.csv>

https://en.wikipedia.org/wiki/Health_effects_of_wine

<https://www.webmd.com/diet/features/health-benefits-wine#1>

<https://www.webmd.com/diet/wine-health-benefits#1>

<https://www.health.com/nutrition/healthiest-wine>

https://www.eatthis.com/drink-wine-every-night/?utm_source=msn&utm_medium=feed&utm_campaign=msn-feed

https://www.eatthis.com/wine-health-benefits/?utm_source=msn&utm_medium=feed&utm_campaign=msn-feed

<https://www.msn.com/en-us/health/nutrition/how-much-wine-is-safe-to-drink-per-day/ar-BB19yG6q>