Wine Tasting with a Sip of Data By: Brandon Jezek, Paula Koziol, Gautam Seenivasan & Jerome Bright



How Do You Pick a Wine?

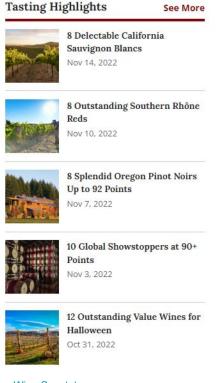
Sort through best sellers lists?



Look for a pretty bottle? Or something on sale?



Look for Featured Wines?



Compare Reviews?

"Ample aromas of blackberry, plum, cherry, licorice, **smoke** & vanilla. Palate shows flavors of ripe fruits with some earth notes of pepper & **cocoa** showing throughout. Finish is long with chocolate & **tobacco** spice..."

"Mulberry, spiced plum, cassis, cinnamon, clove, anise, leather, cream, oak, tobacco, vanilla."

Vivino

"**Medium** plus body. Cherry leather vanilla on the nose. Dark fruit black fruit black cherry on the taste. Pretty balance acidity and tannin. **Pepper** finish"

Wine Spectator

Welcome to Wine-ML! Neural Network Streamlit Natural Random Language Forest Processing



Welcome to Wine-ML!



What qualities are most important for you?

What kind of wines do you like?

Use the input box to describe what kind of wine you're looking for - taste, aroma, pairing, etc and we'll find the best match based on our inventory.

Enter any text information that describes the kind of wine you're looking for, and we'll show you our best match!

dessert riesling

Based on your preferences, we recommend this wine!

Title: BEDELL 2001 LATE HARVEST RIESLING RIESLING (NORTH FORK OF LONG ISLAND)

Variety: RIESLING

Designation: LATE HARVEST RIESLING

Country of origin: US

Province: NEW YORK

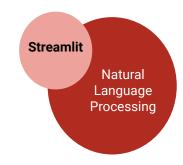
Region: NORTH FORK OF LONG ISLAND, LONG ISLAND

Winery: BEDELL

Taster name: JOE CZERWINSKI

Taster twitter handle: @JOECZ

Description: IT'S PICKED LATE, WHEN THE GRAPES ARE SUPERRIPE, THEN WINEMAKER KIP BEDELL PUT: THEM IN A FREEZER PRIOR TO CRUSH TO FURTHER CONCENTRATE THE RESULTING MUST. THE RESULT I A THICK, GOOEY, UNCTUOUS DESSERT WINE WITH FLAVORS OF TANGERINES AND HONEY. YUM.



Wine Quality Calculator

Use our models to help you pick great quality wines. Our models indicate that alcohol content, volatile acidity, and sulphates are the most important variables when it comes to dictating the quality of a wine.

Don't feel the need to stick to wines that have been reviewed when you can use our models to help you gauge its quality without having to taste it.

Enter as much information that you have about the wine below, and use our models to help you predict its quality!

Fixed acidity

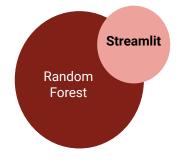
Streamlit

Welcome to Wine-ML!



. . .

The parameters input correspond to a wine with LOW quality (rating of 6 or lower)





Welcome to Wine-ML!



Wine Quality Predictor

Use our models to help you predict the quality of wine based on its reviews!

Enter as much information that you have about the wine below, and use our models to help you predict its quality!

The form is already pre-filled with an example, make sure to DELETE any information that's not relevant for your wine!

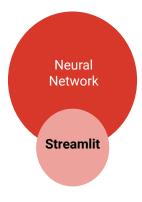
Description

Much like the regular bottling from 2012, this comes across as rather rough and tannic, with rustic, earth

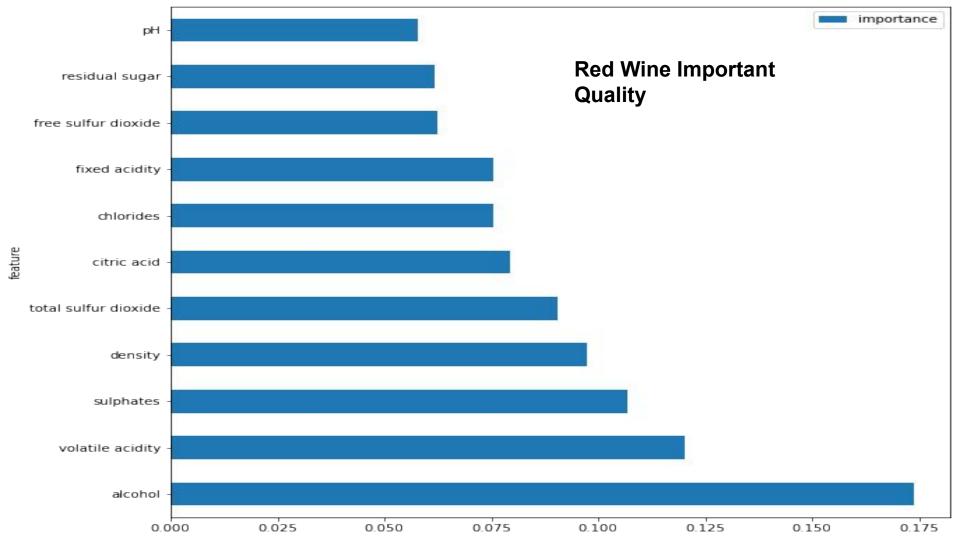
Calculate!

Based on your review, the estimated quality of this wine is:

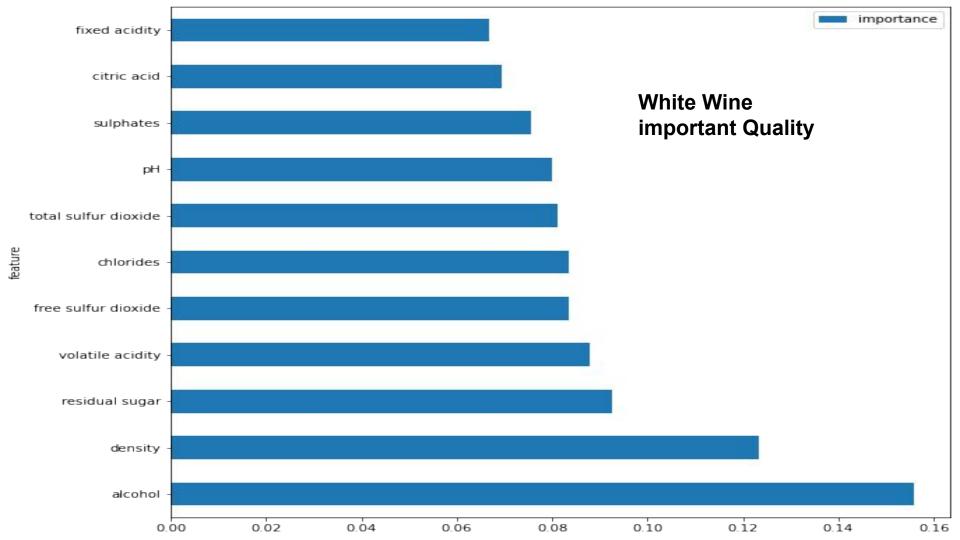
88.79402

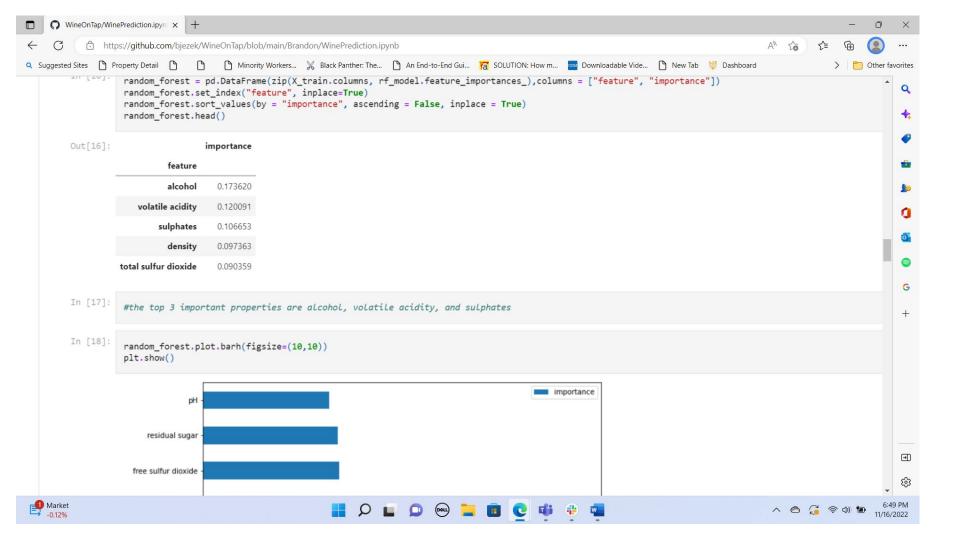


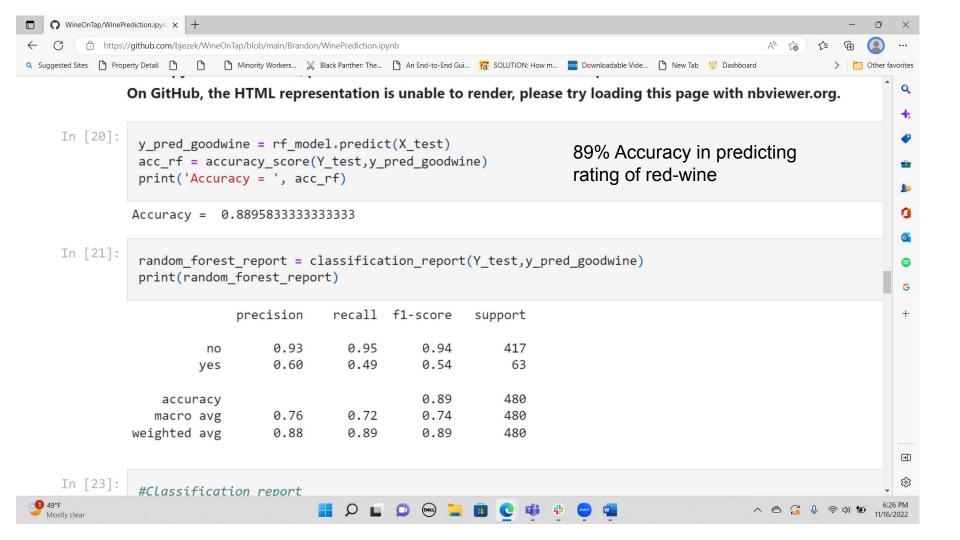
fixed acidity -	1	-0.26	0.67	0.11	0.094	-0.15	-0.11	0.67	-0.68	0.18	-0.062	0.12		-10
volatile acidity -	-0.26			0.0019	0.061	-0.011	0.076	0.022	0.23	-0.26	-0.2	-0.39		- 0.8
citric acid -	0.67	-0.55	1	0.14	0.2	-0.061	0.036	0.36	-0.54	0.31	0.11	0.23		- 0.6
residual sugar -	0.11	0.0019	0.14	ï	0.056	0.19	0.2	0.36	-0.086	0.0055	0.042	0.014		
chlorides -	0.094	0.061	0.2	Red V	Vine In	nporta s	ance N	latrix	-0.27	0.37	-0.22	-0.13		- 0.4
free sulfur dioxide -	-0.15	-0.011	-0.061	o19 Qual	o.oosa lity & Alc	ohol had	o of d the hic	nhest	0.07	0.052	-0.069	-0.051		- 0.2
total sulfur dioxide -	-0.11	0.076	0.036		elation o		1 T	0.071	-0.066	0.043	-0.21	-0.19		
density -	0.67	0.022	0.36	0.36	0.2	-0.022	0.071	1	-0.34	0.15	-0.5	-0.17		- 0.0
рН -	୍-0.68	0.23	-0.54	୍ଦ୍ର.086	-0.27	0.07	40.066	-0.34	1	-0.2	0.21	-0.058		0.2
sulphates -	0.18	-0.26	0.31	0.0055	0.37	0.052	0.043	0.15	-0.2	1	0.094	0.25		0.4
alcohol -	-0.062	-0.2	0.11	0.042	-0.22	-0.069	-0.21	-0.5	0.21	0.094	1	0.48		
quality ∹	0.12		0.23	0.014	-0.13	-0.051	-0.19	-0.17	-0.058	0.25	0.48	1		0.6
1	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	рH	sulphates	alcohol	quality		

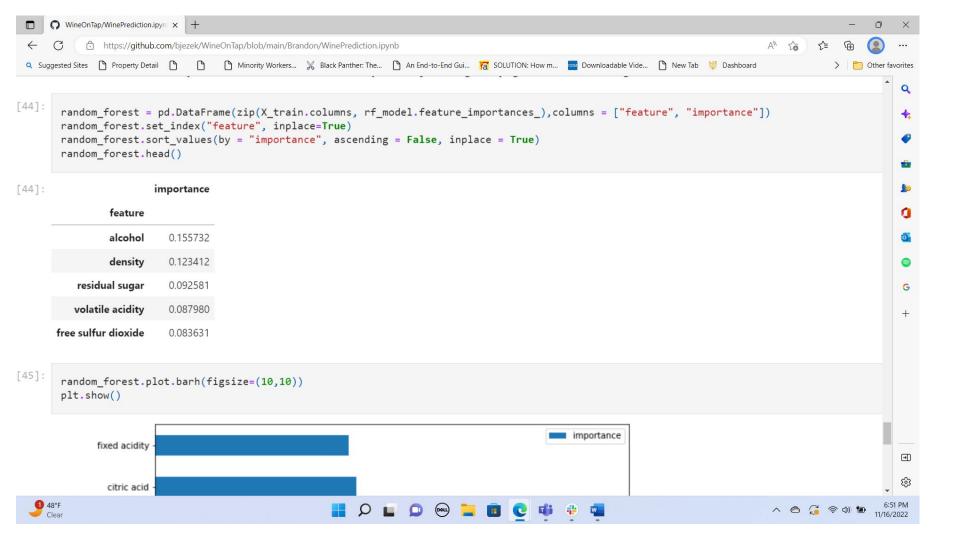


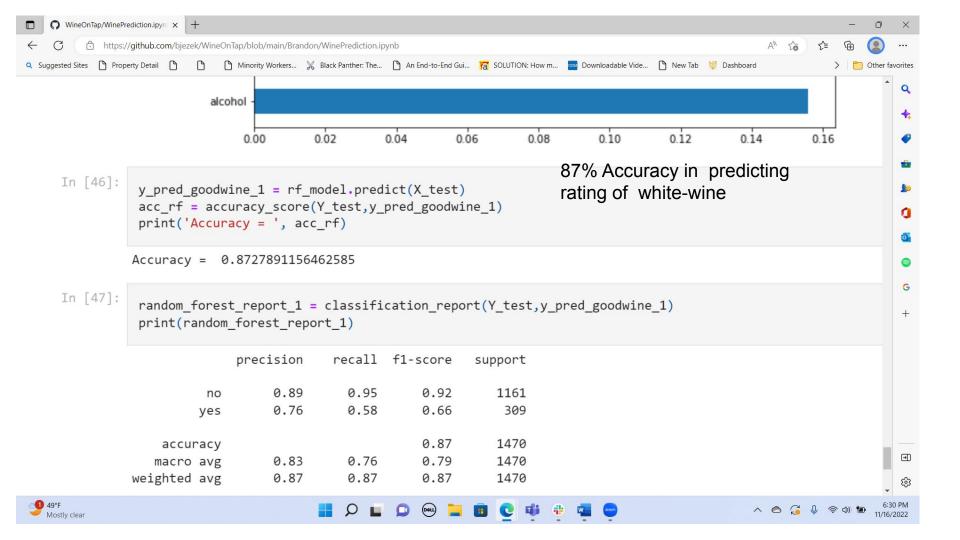
fixed acidity -	1	-0.023	0.29	0.089	0.023	-0.049	0.091	0.27	-0.43	-0.017	-0.12	-0.11	-10	0
volatile acidity -	-0.023	1	-0.15	0.064	0.071	Wh	ite wir	ne limp	ortanc	e Matı	'ix _{0.068}	-0.19	- 0.8	.8
citric acid -	0.29	-0.15	1	0.094	0.11	0.094	0.12	0.15	-0.16	0.062	-0.076	-0.0092	- O.€	.6
residual sugar -	0.089	0.064	0.094	ï	0.089			ine high ol agair			is Qua	lity _{•.098}	- 0.4	
chlorides -	0.023	0.071	0.11	0.089	1	0.1	0.2	0.26	-0.09	0.017	-0.36	-0.21	- 0.4	4
free sulfur dioxide -	-0.049	-0.097	0.094	0.3	0.1	1		0.29	-0.00062	0.059	-0.25	0.0082	- 0.2	.2
total sulfur dioxide -	0.091	0.089	0.12	0.4	0.2			0.53	0.0023	0.13	-0.45	-0.17	- 0.0	.0
density -	0.27	0.027	0.15	0.84	0.26	0.29	0.53	1	-0.094	0.074	-0.78	-0.31		
рн -	-0.43	-0.032	-0.16	-0.19	-0.09	-0.00062	0.0023	-0.094	1	0.16	0.12	0.099	0	-0.2
sulphates -	-0.017	-0.036	0.062	-0.027	0.017	0.059	0.13	0.074	0.16	1	-0.017	0.054	c	0.4
alcohol -	-0.12	0.068	-0.076	-0.45	-0.36	-0.25	-0.45	-0.78	0.12	-0.017	1	0.44	0	-0.6
quality -	-0.11	-0.19	-0.0092	-0.098	-0.21	0.0082	-0.17	-0.31	0.099	0.054	0.44	1		
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality		





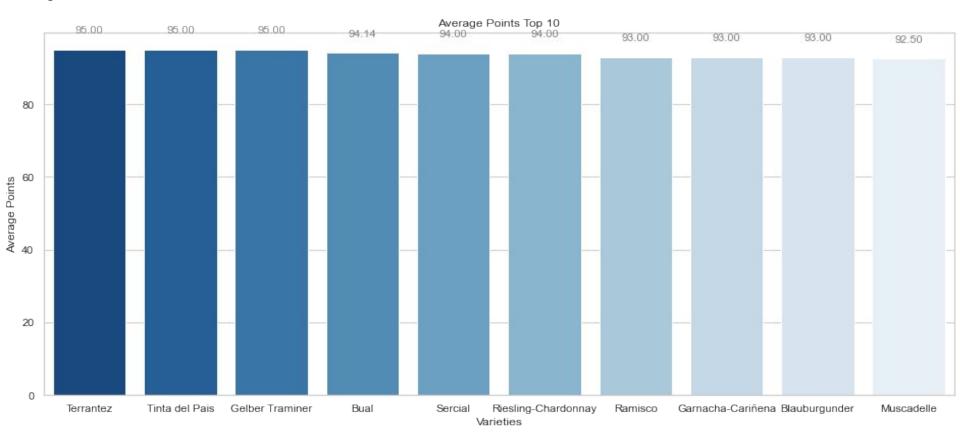




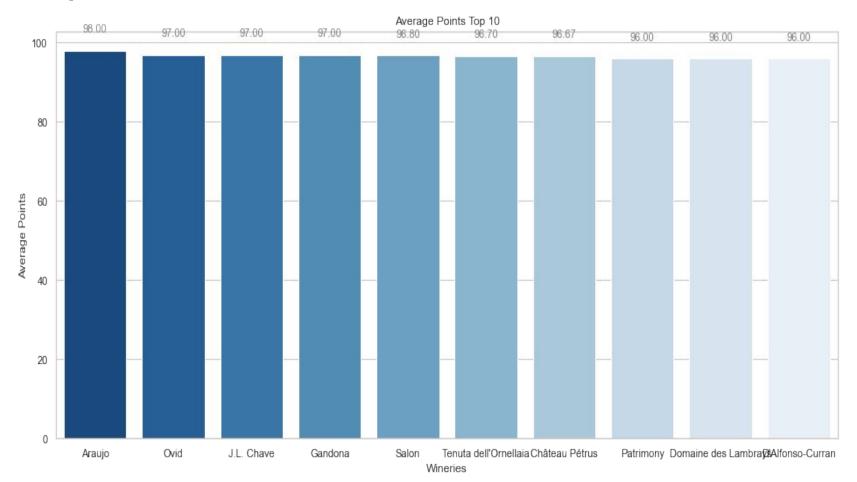


Dataframe Analysis

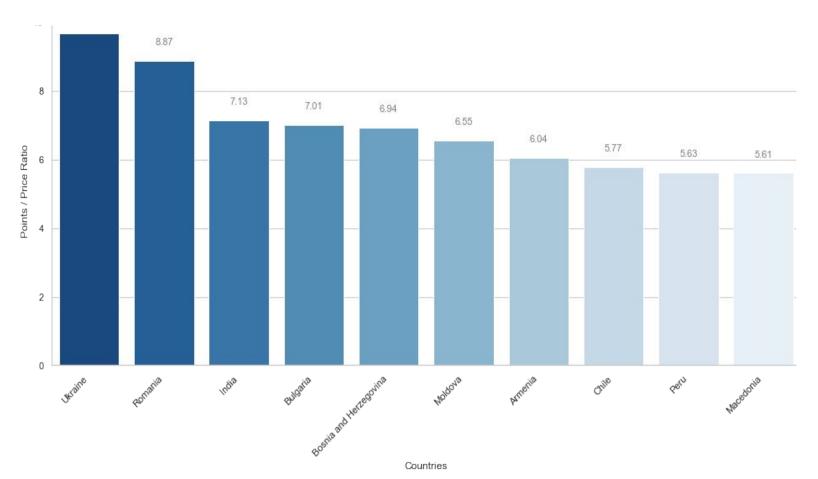
Top 10 Varieties



Top 10 Wineries

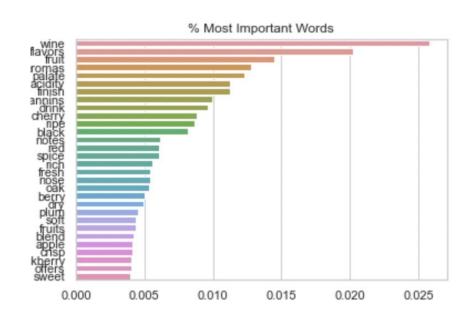


Best Bang for Your Buck



NLP

import seaborn as sns import matplotlib.pyplot as plt from nltk.tokenize.treebank import TreebankWordDetokenizer from nltk.corpus import stopwords from nltk import word_tokenize



Sentiment Analysis

import nltk

from nltk.sentiment.vader import SentimentIntensityAnalyzer

import nltk

nltk.download('vader lexicon')

```
[156]: df sentiment = pd.read csv('winemag-data-130k-v2.csv')
  []: #Sentiment analysis of wine reviews
[157]: sid = SentimentIntensityAnalyzer()
       df_sentiment.reset index(inplace=True, drop=True)
       df_sentiment[['neg', 'neu', 'pos', 'compound']] = df_sentiment['description'].apply(sid.polarity_scores).apply(pd.Series)
       df sentiment
       df_sentiment["variety"] = df["variety"]
       df sentiment.head()
      lesignation points price province region_1 region_2 taster_name taster_twitter_handle
                                                                                                     title
                                                                                                             variety
                                                                                                                                                  compound
                                                                                                  Nicosia
                                                                                                              White
            Vulkà
                                                                    Kerin
                                                                                               2013 Vulkà
                                 Sicily &
                                                                                  @kerinokeefe
                                                                                                                       Nicosia 0.000 0.935 0.065
                                                        NaN
                                                                                                                                                      0.1531
           Bianco
                                                                  O'Keefe
                                                                                                   Bianco
                                                                                                              Blend
                                                                                                   (Etna)
                                                                                                Quinta dos
                                                                                                Avidagos
                                                                                                                        Quinta
                                                                                                    2011
                                                                                                         Portuguese
                                                                                                                          dos 0.000 0.868 0.132
                                                                                                                                                      0.6486
         Avidagos
                      87
                          15.0
                                  Douro
                                              NaN
                                                        NaN
                                                               Roger Voss
                                                                                   @vossroger
                                                                                                 Avidagos
                                                                                                                      Avidagos
                                                                                                     Red
                                                                                                  (Douro)
                                                                                                Rainstorm
                                                                                                2013 Pinot
             NaN
                          14.0
                                                              Paul Gregutt
                                                                                                     Gris
                                                                                                           Pinot Gris Rainstorm 0.053 0.947 0.000
                                                                                                                                                     -0.1280
                                                                                   @paulgwine
                                                                                               (Willamette
                                                                                                  Valley)
                                                                                                St. Julian
                                                                                                    2013
          Reserve
                                             Lake
                                                                 Alexander
                                                                                                  Reserve
             Late
                      87
                         13.0 Michigan
                                          Michigan
                                                        NaN
                                                                                         NaN
                                                                                                             Riesling St. Julian 0.000 0.926 0.074
                                                                                                                                                      0.3400
                                                                                                     Late
                                                                  Peartree
                                            Shore
          Harvest
                           Screenshot
```

Filtering out for High Sentiment

```
filter5 = df_palate["Sentiment Score"] > .9
good_palate = df_palate.loc[filter5]
good_palate.dropna()
```

	description	Sentiment Score	Variety
119	Medium-gold in color. Complex and inviting nos	0.9732	Riesling
128	Compelling minerality on the nose, Refined and	0.9313	Pinot Blanc
133	Einaudi's wines have been improving lately, an	0.9325	Nebbiolo
135	The color is just beginning to show signs of b	0.9398	Nebbiolo
147	Black cherry, black plum and black currant are	0.9386	Cabernet Sauvignon
•••		***	
129907	Toasty aroams of grilled apple, peach and hone	0.9322	Chardonnay
129908	Pleasant scents of toasted apple and peach pit	0.9528	Chardonnay
129917	This dark, meaty Malbec has some Agrelo fruit	0.9413	Malbec
129921	There is a select group of under-\$20 Malbecs f	0.9231	Malbec
129930	Arguably on the heavy and rich side, but if yo	0.9387	Tempranillo Blend

Final Recommendation

Streamlit Integration

Make a selection below to find the best wines associated with your preference:







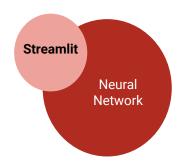
Neural networks are useful in predictive analysis because of the model's hidden layers.

The neural network uses the hidden layer to make predictions with higher precision because learns in similar way to humans. That's because it 'learns' the way a human does

Wine reviews are intended for consumers who want to make the best decisions providing them with the most descriptive information available but in many instances consumers gain more information on different wines and enable them to select wines to their liking more easily.

Using a Neural network model, the goal is to produce information that customers can use to make a good decision about wines before making their purchases.

Objective: Train Neural Network model predict to predict wine scores based on their reviews



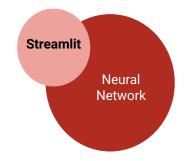




Observe the overal data distribution

wine_data.describe()

	Unnamed: 0	points	price
count	150930.000000	150930.000000	137235,000000
mean	75464.500000	87.888418	33.131482
std	43569.882402	3.222392	36.322536
min	0.000000	80.000000	4.000000
25%	37732.250000	86.000000	16.000000
50%	75464.500000	88.000000	24.000000
75%	113196.750000	90.000000	40.000000
max	150929.000000	100.000000	2300.000000





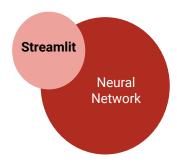
Welcome to Wine-ML!

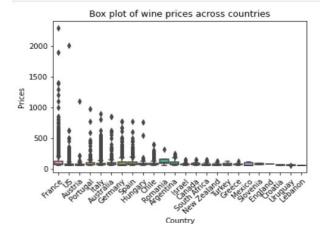


Prices vary by countries

```
#Extract the most expensive wines
priceVScountry = wine_data[["price", "country"]].dropna(how = "any").nlargest(20000, "price")

ax = sns.boxplot(x="country", y = "price", data=priceVScountry)
ax.set_title("Box plot of wine prices across countries")
ax.set(xlabel = "Country", ylabel = "Prices")
ax.set_xticklabels(ax.get_xticklabels(),rotation=45,ha="right",rotation_mode='anchor')
plt.show()
```

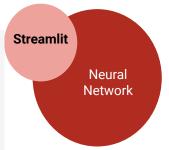






Welcome to Wine-ML!







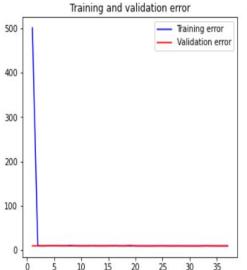


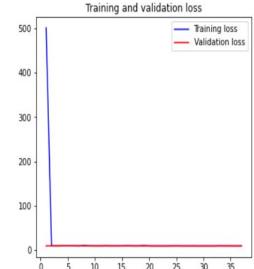
Welcome to Wine-ML!

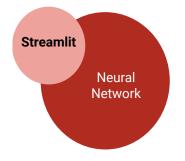


Epoch 00037: val_loss did not improve from 9.48354

Training error: 9.5225 Testing error: 9.7543







Expenses

Finalizing the development costs:

- 1) Analysis and raw estimation Free
- 2) Prototype development \$5000
- 3) MVP development (prototype included) \$10000
- 4) Deployment and Release \$5000

How Much Does It Cost To Build A Recommendation System (azati.ai)

ROI and Pricing

- Charge a flat \$10,000 a year fee to recommend wines
- Charge a maintenance fee of \$250 a month
- Include a affiliate performance clause for a 3% royalty after a certain milestone is reached.

• 50%+ ROI per client

Next Steps





Our Deep Learning models can produce improved results with more information and stringent model tuning.

The benefits of these improved tests would allow us to explore how these models can assess consumer reviews and surveys. These studies could provide the customer with earlier consumer feedback as compared to the typical wine review of the professionals, like movie scores.

Additionally, we can apply these models to consumer tastes and expectations for vodka, tequila, or beer and supply similar benefits for customers as they try to gather the information they consider important for their next purchase. &

Run over & Undersampling check hedging parameters for accuracy results on a continued basis