Wine Analysis
SPIRITS BEER & MORE

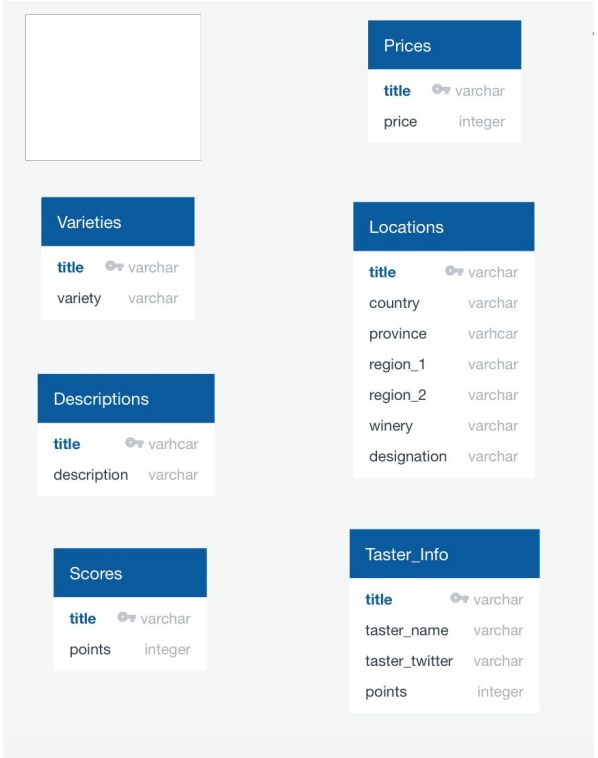
Topic

We are a team within a startup wine and spirits company that have been tasked with assisting the Operations Team in selecting which wine to increase production on, and determining the association of reviews and descriptors from data scraped from WineEnthusiast.com during the week of November 22nd, 2017. This topic was selected as we all wanted to work with a product or data type that we were either familiar with or would be interested in learning more about

Data

- Points: the number of points WineEnthusiast rated the wine on a scale of 1-100 (though they say they only post reviews for wines that score >=80)
- Title: the title of the wine review, which often contains the vintage if you're interested in extracting that feature
- Variety: the type of grapes used to make the wine (ie Pinot Noir)
- Description: a few sentences from a sommelier describing the wine's taste, smell, look, feel, etc.
- Country: the country that the wine is from
  Province: the province or state that the wine is from
  Region 1: the wine growing area in a province or state (ie Napa)
  Region 2: sometimes there are more specific regions specified within a wine growing area (ie Rutherford inside the Napa Valley), but this value can sometimes be blank
  Winery: the winery that made the wine
  Designation: the vineyard within the winery where the grapes that made the wine are from
  Price: the cost for a bottle of the wine
  Taster Name: name of the person who tasted and reviewed the wine
  Taster Twitter Handle: Twitter handle for the person who tasted and reviewed the wine

Data Mappings



Wine by Description

title	description
Nicosia 2013 Vulkv† Bianco (Etna)	Aromas include tropical fruit, broom, brimstone and dried herb. Th
Quinta dos Avidagos 2011 Avidagos Red (Douro)	This is ripe and fruity, a wine that is smooth while still structured.
Rainstorm 2013 Pinot Gris (Willamette Valley)	Tart and snappy, the flavors of lime flesh and rind dominate. Some
St. Julian 2013 Reserve Late Harvest Riesling (Lake Michigan S	Pineapple rind, lemon pith and orange blossom start off the aroma
Sweet Cheeks 2012 Vintner's Reserve Wild Child Block Pinot N	Much like the regular bottling from 2012, this comes across as rat
Tandem 2011 Ars In Vitro Tempranillo-Merlot (Navarra)	Blackberry and raspberry aromas show a typical Navarran whiff of
Terre di Giurfo 2013 Belsito Frappato (Vittoria)	Here''s a bright, informal red that opens with aromas of candied b
Trimbach 2012 Gewurztraminer (Alsace)	This dry and restrained wine offers spice in profusion. Balanced wi
Heinz Eifel 2013 Shine Gewvºrztraminer (Rheinhessen)	Savory dried thyme notes accent sunnier flavors of preserved peac

Wine by Location

title	cou	province	region_1	region_2	winery	designation	
Nicosia 2013 Vulkv† Bianco (Etna)	Italy	Sicily & Sardinia	Etna		Nicosia	Vulkv† Bianco	
Quinta dos Avidagos 2011 Avidagos Red (Douro)	Portugal	Douro			Quinta dos Avidagos	Avidagos	
Rainstorm 2013 Pinot Gris (Willamette Valley)	US	Oregon	Willamette Valley	Willamette Valley	Rainstorm		
St. Julian 2013 Reserve Late Harvest Riesling (Lake	US	Michigan	Lake Michigan Shore		St. Julian	Reserve Late Harvest	
Sweet Cheeks 2012 Vintner's Reserve Wild Child Bl	US	Oregon	Willamette Valley	Willamette Valley	Sweet Cheeks	Vintner's Reserve Wild Child Block	
Tandem 2011 Ars In Vitro Tempranillo-Merlot (Nav	Spain	Northern Spain	Navarra		Tandem	Ars In Vitro	
Terre di Giurfo 2013 Belsito Frappato (Vittoria)	Italy	Sicily & Sardinia	Vittoria		Terre di Giurfo	Belsito	
Trimbach 2012 Gewurztraminer (Alsace)	France	Alsace	Alsace		Trimbach		
Heinz Eifel 2013 Shine GewVºrztraminer (Rheinhes	Germany	Rheinhessen			Heinz Eifel	Shine	

Wine by Price

title	price
Nicosia 2013 VulkV† Bianco (Etna)	
Quinta dos Avidagos 2011 Avidagos Red (Douro)	15
Rainstorm 2013 Pinot Gris (Willamette Valley)	14
St. Julian 2013 Reserve Late Harvest Riesling (Lake Micl	13
Sweet Cheeks 2012 Vintner's Reserve Wild Child Block I	65
Tandem 2011 Ars In Vitro Tempranillo-Merlot (Navarra	15
Terre di Giurfo 2013 Belsito Frappato (Vittoria)	16
Trimbach 2012 Gewurztraminer (Alsace)	24
Heinz Eifel 2013 Shine GewVºrztraminer (Rheinhessen)	12

Wine by Score

title	points
Nicosia 2013 Vulkv† Bianco (Etna)	87
Quinta dos Avidagos 2011 Avidagos Red (Douro)	87
Rainstorm 2013 Pinot Gris (Willamette Valley)	87
St. Julian 2013 Reserve Late Harvest Riesling (Lake Michigan S	87
Sweet Cheeks 2012 Vintner's Reserve Wild Child Block Pinot N	87
Tandem 2011 Ars In Vitro Tempranillo-Merlot (Navarra)	87
Terre di Giurfo 2013 Belsito Frappato (Vittoria)	87
Trimbach 2012 Gewurztraminer (Alsace)	87
Heinz Eifel 2013 Shine GewvºrztraminerÂ (Rheinhessen)	87

Wine by Taster Info

title	taster_name	taster_twitter	points
Nicosia 2013 Vulkv† Bianco (Etna)	Kerin O,ÄôKeefe	@kerinokeefe	87
Quinta dos Avidagos 2011 Avidagos Red (	Roger Voss	@vossroger	87
Rainstorm 2013 Pinot Gris (Willamette V	Paul Gregutt	@paulgwine¬†	87
St. Julian 2013 Reserve Late Harvest Ries	Alexander Peartree		87
Sweet Cheeks 2012 Vintner's Reserve Wi	Paul Gregutt	@paulgwine¬†	87
Tandem 2011 Ars In Vitro Tempranillo-M	Michael Schachner	@wineschach	87
Terre di Giurfo 2013 Belsito Frappato (Vit	Kerin O,ÄôKeefe	@kerinokeefe	87
Trimbach 2012 Gewurztraminer (Alsace)	Roger Voss	@vossroger	87
Heinz Eifel 2013 Shine GewVºrztraminer	Anna Lee C. Iijima		87

Wine by Variety

title	variety
Nicosia 2013 Vulkv† Bianco (Etna)	White Blend
Quinta dos Avidagos 2011 Avidagos Red (Douro)	Portuguese Red
Rainstorm 2013 Pinot Gris (Willamette Valley)	Pinot Gris
St. Julian 2013 Reserve Late Harvest Riesling (Lake Michigan Sl	Riesling
Sweet Cheeks 2012 Vintner's Reserve Wild Child Block Pinot Nc	Pinot Noir
Tandem 2011 Ars In Vitro Tempranillo-Merlot (Navarra)	Tempranillo-Merlot
Terre di Giurfo 2013 Belsito Frappato (Vittoria)	Frappato
Trimbach 2012 Gewurztraminer (Alsace)	GewV⁰rztraminer
Heinz Eifel 2013 Shine GewV⁰rztraminer (Rheinhessen)	GewV⁰rztraminer

Questions

1. Which wines should we choose to produce more of based on the reviews of our consumers
2. Which winery should we choose to work with based on the variety we select and the reviews provided?
3. What is the average price of wine receiving above average reviews?

Data Analysis

Cleaning the Data

In [5]:
```
#drop description column as it does not apply
wine_df = wine_df.drop(columns=['description'])
```

In [6]:
```
#check for null values in the points column
wine_df_check = pd.isnull(wine_df["points"])
```

In [7]:
```
wine_df_check
```

Out[7]:
```
0          False
1          False
2          False
3          False
4          False
           ...
150925     False
150926     False
150927     False
150928     False
150929     False
Name: points, Length: 150930, dtype: bool
```

In [8]:
```
#drop rows from points column without points
#wine_df['points'] = wine_df['points'].dropna
```

In [9]:
```
#fill null variety column with other
wine_df['variety'] = wine_df['variety'].fillna('Other')
```

In [10]:
```
#fill null winery column with none
wine_df['winery'] = wine_df['winery'].fillna('None')
```

In [11]:
```
#read df
wine_df.head()
```

Out[11]:

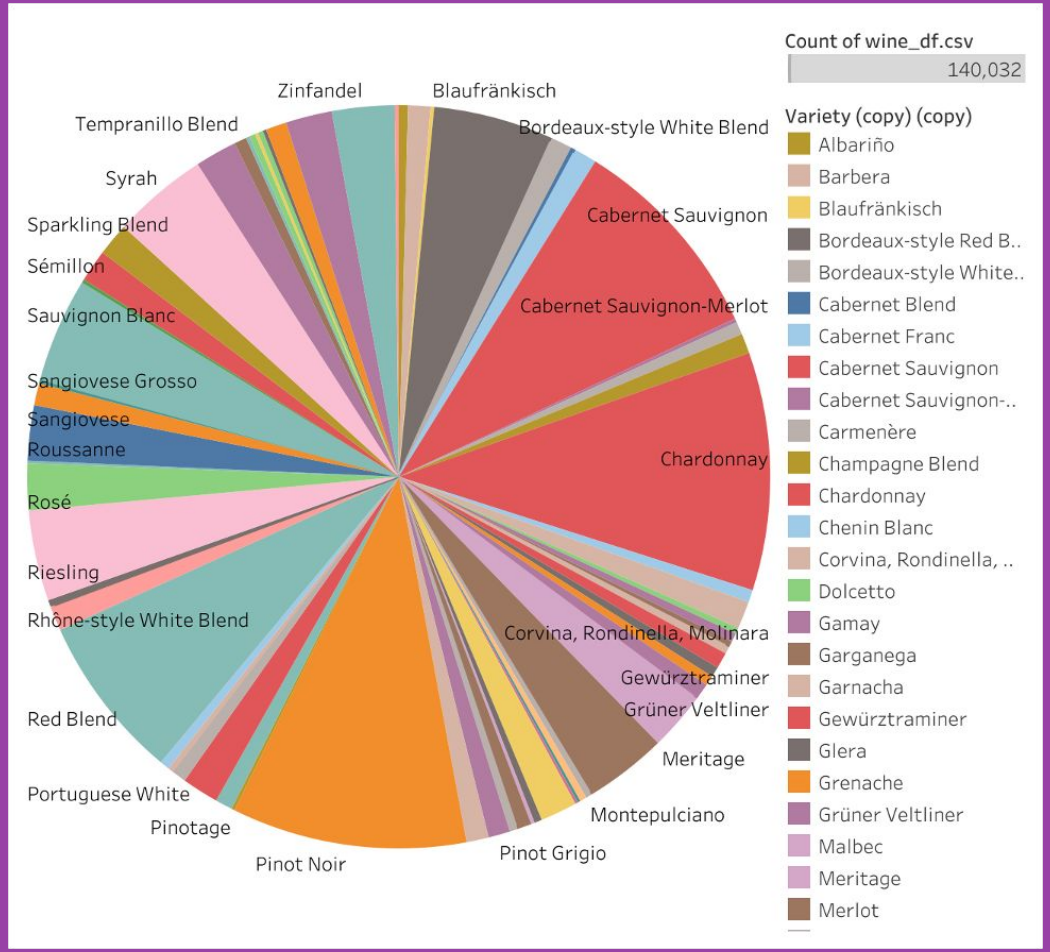
	country	designation	points	price	province	region_1	region_2	variety	winery
0	US	Martha's Vineyard	96	235.0	California	Napa Valley	Napa	Cabernet Sauvignon	Heitz
1	Spain	Carodorum Selección Especial Reserva	96	110.0	Northern Spain	Toro	NaN	Tinta de Toro	Bodega Carmen Rodríguez
2	US	Special Selected Late Harvest	96	90.0	California	Knights Valley	Sonoma	Sauvignon Blanc	Macauley
3	US	Reserve	96	65.0	Oregon	Willamette Valley	Willamette Valley	Pinot Noir	Ponzi
4	France	La Brûlade	95	66.0	Provence	Bandol	NaN	Provence red blend	Domaine de la Bégude

:
```
#remove any points lower than 85
variety_filter_df = wine_df.groupby(wine_df["points"] > 90)
```

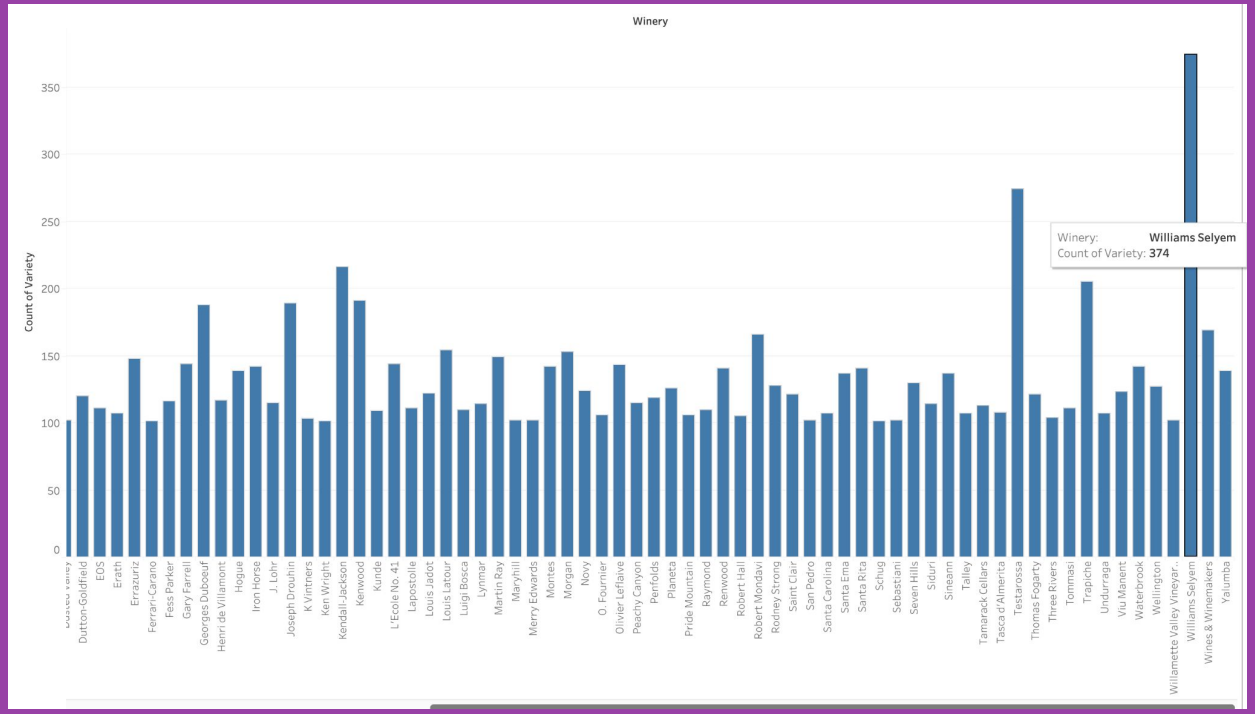
Findings

```
#get counts of varieties with the highest number of high ratings
variety_counts = wine_df["variety"].value_counts(ascending=False)
variety_counts
```

```
Chardonnay                  14482
Pinot Noir                  14291
Cabernet Sauvignon          12800
Red Blend                   10062
Bordeaux-style Red Blend     7347
                            ...
Carignan-Syrah                  1
Premsal                         1
Muskat                          1
Syrah-Carignan                  1
Carnelian                       1
Name: variety, Length: 632, dtype: int64
```



Findings




```
#get counts of the winery values
winery_counts= winery_filter_df["winery"].value_counts(ascending=False)
winery_counts
```

```
variety  winery
False    Williams Selyem          374
         Testarossa               274
         DFJ Vinhos               258
         Chateau Ste. Michelle    225
         Columbia Crest           217
                                 ...
         l'Escargot                 1
         l'homme qui ris            1
         the Ghost of 413           1
         Ébano                      1
         áster                      1
Name: winery, Length: 14810, dtype: int64
```

Findings

In [16]:
```
#get average pricing of the selected varieties
average_filter_df = wine_df.groupby(wine_df["variety"] == ("Chardonnay","Cabernet Sauvignon", "Pinot Noir")).mean()
average_filter_df
```

Out[16]:

variety	points	price
False	87.888418	33.131482

Predictions

1. Can we predict if consumers will purchase the type of wine we choose to produce based on their reviews?
2. Can we predict the scores of future wine tastings based on previous ones?
3. Does the description of the wines lead to a higher reviews/points?

Machine Learning

- On the Machine Learning part, we applied the Keras model from Tensorflow. In the first data run, the columns “Description” and “” were removed from the algorithm, while a second run was made including the “Description” column in order to compare values between them to see which model gives more accuracy.

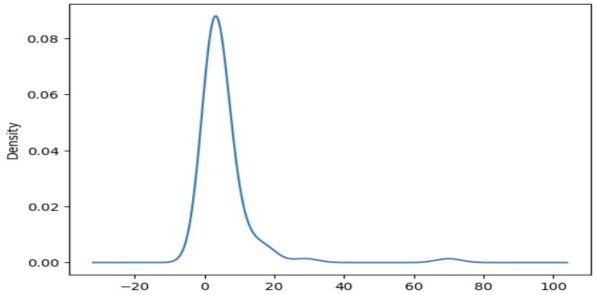
- In order to carry out the ML model applied to the Wine_Analysis dataset, all the columns, except “Designation”, were cleaned, transformed and encoded. Later on, Standard Scaler was applied for normalizing the features, and all data frame were merged into the original.

Machine Learning: Data Cleaning

```
In [5]: region_counts = wine_df.region_1.value_counts()
        region_counts
```

```
Out[5]: Napa Valley                          6246
        Columbia Valley (WA)                4994
        Mendoza                             3597
        Russian River Valley                3588
        California                          3472
                                            ...
        Vin de Pays des Côtes de Thongue       1
        Toledo                                 1
        Vin de Pays de Montferrand             1
        Napa County-Lake County                1
        El Pomar District                      1
        Name: region_1, Length: 1239, dtype: int64
```

click to scroll output; double click to hide




```
In [21]: word_list = ['full bodied','earthy','savory','herbaceous','minerally','toasty','oaky',
                      'spicy','flabby','supple','balanced','sweet','aerated','buttery','fruity']
```

```
In [22]: def check_words(row):
             for word in word_list:
                 if word in row:
                     return word
             return 'none'
```

```
In [23]: wine_df['adjectives'] = wine_df['description'].apply(check_words)
```

```
In [7]: def get_sentiment(review):
            blob = TextBlob(review)
            return blob.sentiment.polarity
```

```
In [8]: wine_df['sentiment'] = wine_df['description'].apply(get_sentiment)
```

Machine Learning: Creating The Model

```
number_input_features = len(X_train_scaled[0])
hidden_layers_1 = 750
hidden_layers_list = [250]

nn = tf.keras.models.Sequential()

nn.add(tf.keras.layers.Dense(units=hidden_layers_1, input_dim=number_input_features,
                             activation='relu',
                             kernel_regularizer=regularizers.l1_l2(l1=0.01, l2=0.01)))
nn.add(tf.keras.layers.Dropout(0.5))
for i in hidden_layers_list:
    nn.add(tf.keras.layers.Dense(units=i, activation='relu',
                                 kernel_regularizer=regularizers.l1_l2(l1=0.01, l2=0.01)))
    nn.add(tf.keras.layers.Dropout(0.5))

nn.add(tf.keras.layers.Dense(units=1, activation='relu'))
```

In [40]:
```
early_stop = EarlyStopping(monitor='val_mae',patience = 5, verbose = 1)
```

In [41]:
```
start_time = time.time()
for i in range(1,82):
    if i >= 60 and i < 72:
        early_stop = EarlyStopping(monitor='val_mae',patience = 3, verbose = 1)
    elif i >= 72:
        early_stop = EarlyStopping(monitor='val_mae',patience = 1, verbose = 1)
    nn.fit(X_train_scaled,y_train,epochs=10000,validation_data=(X_test,y_test),callbacks=
    if i % 2 != 0:
        print(f"loaded Weights {i}")
        nn.load_weights('best_weights.h5')
    else:
        print(f"Random Weights {i}")

end_time = time.time()
```

Findings

```
In [45]: loss,accuracy = model_best.evaluate(X_test,y_test)
         13/13 [==============================] - 0s 3ms/step - loss: 10.5437 - mae: 1.9130
```

```
In [54]: pred_df
```

Out[54]:

	country	price	province	region_1	region_2	variety	winery	sentiment
0	US	30	California	Napa Valley	Napa	Chardonnay	Hall	2.240741

```
In [65]: y_pred_df
```

Out[65]:

	prediction
0	87.142853

Machine Learning_Sentiment Analysis

```
#Creating the sentiment column to store the values under condition.
wine_df['sentiment'] = wine_df['points'].apply(lambda x: 1. if x >= 90 else 0.)
```

```
wine_df = wine_df[['description','price','points','sentiment']]
wine_df
```

Out[6]:

	description	price	points	sentiment
0	This tremendous 100% varietal wine hails from ...	235.0	96	1.0
1	Ripe aromas of fig, blackberry and cassis are ...	110.0	96	1.0
2	Mac Watson honors the memory of a wine once ma...	90.0	96	1.0
3	This spent 20 months in 30% new French oak, an...	65.0	96	1.0
4	This is the top wine from La Bégude, named aft...	66.0	95	1.0
...	...	...	...	...
150925	Many people feel Fiano represents southern Ita...	20.0	91	1.0
150926	Offers an intriguing nose with ginger, lime an...	27.0	91	1.0
150927	This classic example comes from a cru vineyard...	20.0	91	1.0
150928	A perfect salmon shade, with scents of peaches...	52.0	90	1.0
150929	More Pinot Grigios should taste like this. A r...	15.0	90	1.0

150930 rows × 4 columns

In [22]:
```
# Evaluate the model using the test data
test_loss, test_acc = model.evaluate(testing_padded, testing_labels_final)
print('Test Loss:', test_loss)
print('Test Accuracy:', test_acc)
```

```
944/944 [==============================] - 1s 1ms/step - loss: 0.2560 - accuracy: 0.9053
Test Loss: 0.25595954060554504
Test Accuracy: 0.9052872061729431
```

In [23]:
```
#Creating a confusion matrix
#Get the probability scores for each class,
#Get the index of the class with the highest probability
y_pred_probs = model.predict(testing_padded)
y_pred = np.argmax(y_pred_probs, axis=1)
conf_matrix = confusion_matrix(testing_labels_final, y_pred)
conf_matrix

```

```
944/944 [==============================] - 1s 1ms/step
```

Out[23]:
```
array([[21267,     0],
       [ 8919,     0]], dtype=int64)
```

Wine Selector Website

An additional aspect of this project was to create a website using JavaScript and HTML that allows users to filter through the WineEnthusiast data and select a wine according to their specific tastes. You are able to select some or all of the categories and type whatever you like, and the selector tool will draw from the dataset and show you all or any of the wines that match your search.

Wine Recommendation Selector

Which Wine is Right For You?

Wine Selector Tool

Personalized wine recommendations we know you'll love

Our wine selector tool is a quick and easy-to-use system that narrows down your perfect pick based on the kinds of flavors you like, your budget, and winery location.

Data about the wines were scraped from articles and reviews from WineEnthusiast magazine during the week of November 22, 2017, so the filter also gives you information about the wine's score, reviewer, and reviewer's Twitter handle.

Use this selector to get wine suggestions specifically catered to your tastes. Or use it to remind you of that delicious wine you recently enjoyed.

FILTER SEARCH

Enter Variety

Cabernet Sauvignon

Enter Country

US

Enter Province

California

Enter Region 1

Napa Valley

Enter Region 2

Napa

Enter Winery

Okapi

Enter Designation

Estate

Enter Price

100

Filter Table

Clear Table

Score	Title	Description	Taster Name	Taster Twitter
87	Nicosia 2013 Vulkà Bianco (Etna)	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.	Kerin O'Keefe	@kerinokeefe
87	Quinta dos Avidagos 2011 Avidagos Red (Douro)	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.	Roger Voss	@vossroger
87	Rainstorm 2013 Pinot Gris (Willamette Valley)	Tart and snappy, the flavors of lime flesh and rind dominate. Some green pineapple pokes through, with crisp acidity underscoring the flavors.	Paul Gregutt	@paulgwine

Price	Designation	Variety	Region 1	Region 2	Province	Country	Winery
	Vulkà Bianco	White Blend	Etna		Sicily & Sardinia	Italy	Nicosia
15	Avidagos	Portuguese Red			Douro	Portugal	Quinta dos Avidagos
14		Pinot Gris	Willamette Valley	Willamette Valley	Oregon	US	Rainstorm