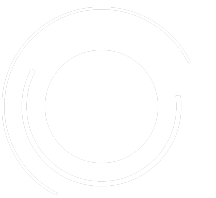


04-2020

AIDI 2004 - AI IN ENTERPRICE FINAL PROJECT

DANIEL ADEGOKE

100711738





WINE RECOMMENDER SYSTEM

WHAT IT DOES

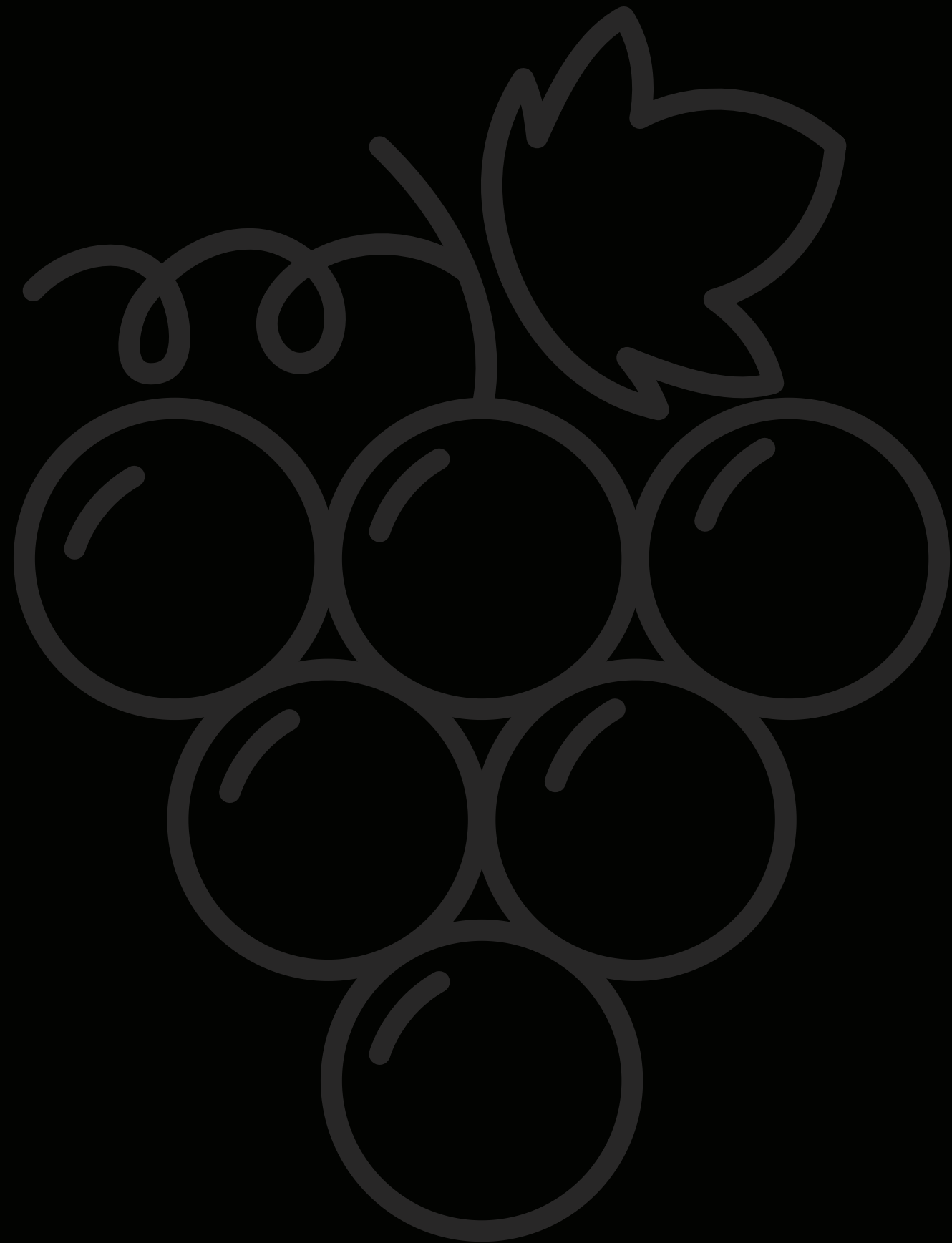
Its an **AI system** that can comb through a large dataset of expert wine reviews, analyze the data, and compute a shortlist of **recommendations.**



WHO ITS FOR

This is for new **wine drinkers** who don't know how or what flavour to choice when getting wine.





HOW ITS DONE

Limit our machine learning algorithm to the top six varieties of wine among the entire dataset

PROJECT OUTCOME

MINIMIZE THE USER'S SEARCH,
AND RECOMMENDED WINE(S)
FROM THE DATA.



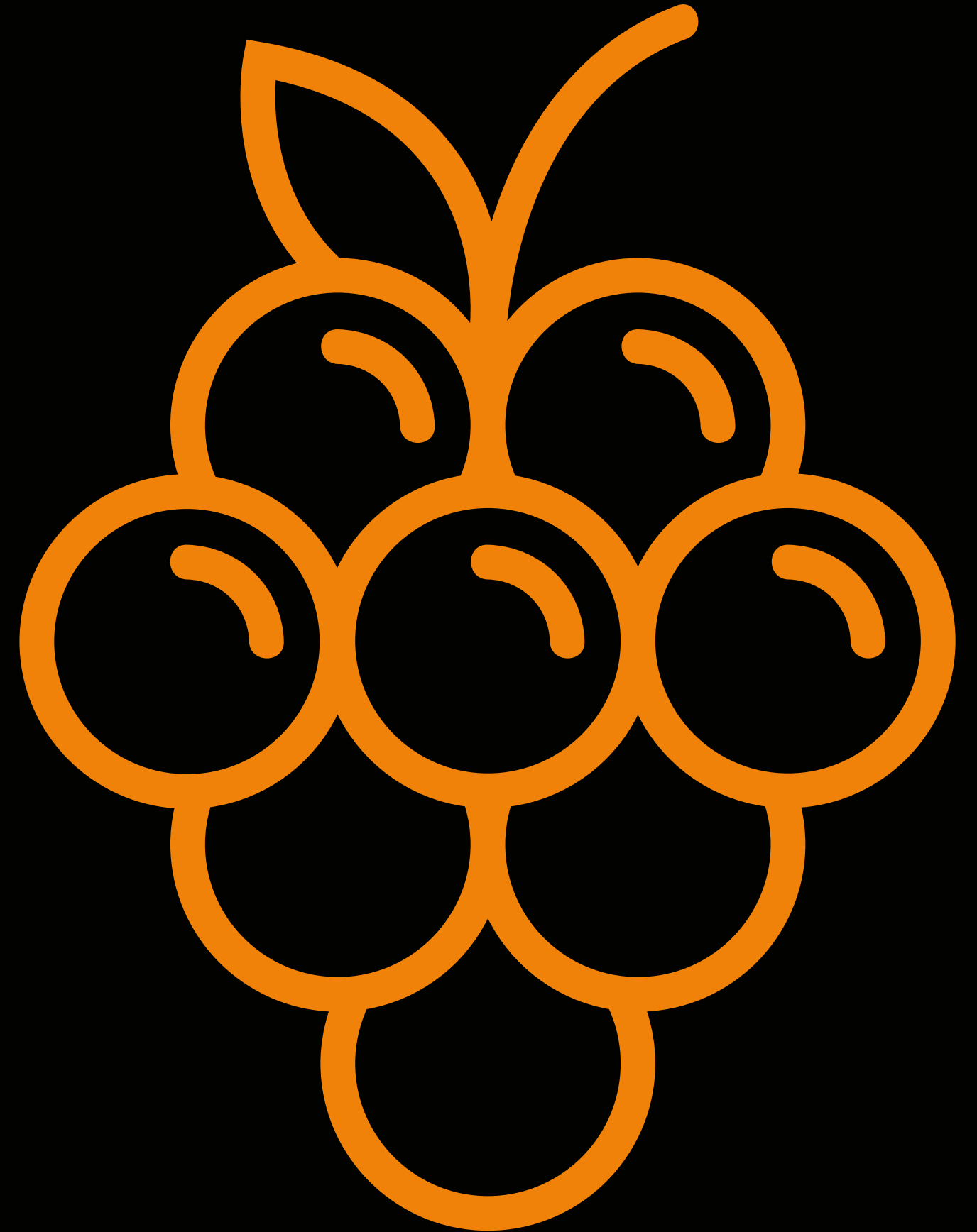
DATA

The dataset used is a public dataset that contains around ~130000 wine reviews from the WineEnthusiast website, which is a community of wine experts

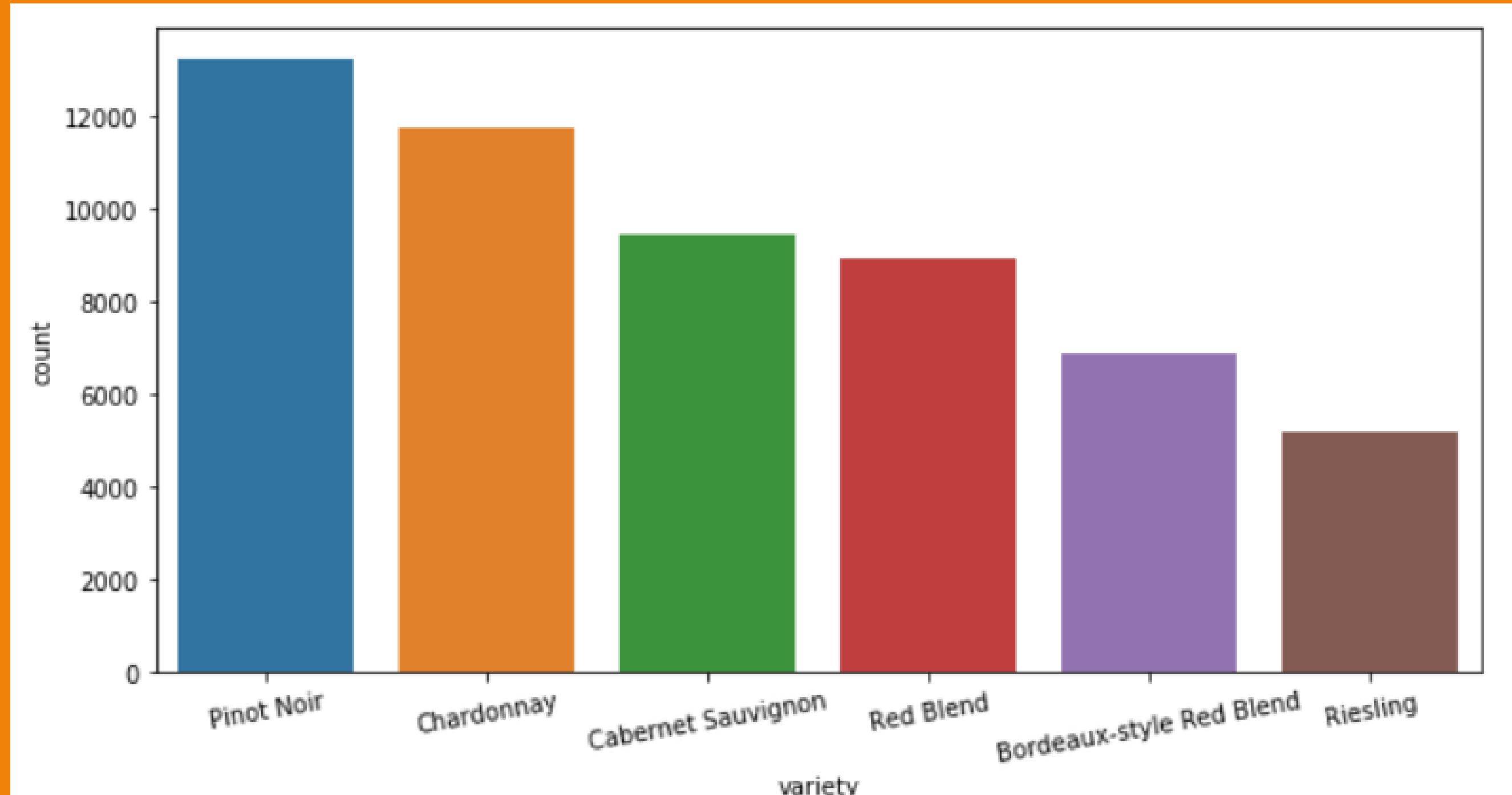


For this dataset, we have **707 different grapes**. After getting the recommend grape/variety we will list the top wine titles of that grape using the points and price features of our dataset.

The constraint with this approach is that some of the most known grapes have thousands of more observations than the not known grapes. In consequence, we have a huge **imbalanced class** problem.



Top six grape varieties based on the number of observations.



SMOTE will be used for imbalancing problem

tf-idf for text vectorization due to size of text data and difficulty

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 129971 entries, 0 to 129970
Data columns (total 13 columns):
country                129908 non-null object
description            129971 non-null object
designation            92506 non-null object
points                129971 non-null int64
price                 120975 non-null float64
province              129908 non-null object
region_1              108724 non-null object
region_2              50511 non-null object
taster_name          103727 non-null object
taster_twitter_handle 98758 non-null object
title                 129971 non-null object
variety               129970 non-null object
winery                129971 non-null object
dtypes: float64(1), int64(1), object(11)
memory usage: 12.9+ MB
```

1 Type Float(Price) and 1 Integer (points), the rest are text/object

```
df.shape
```

```
(129971, 13)
```

Almost 130000 observations of different types of wine and 13 features

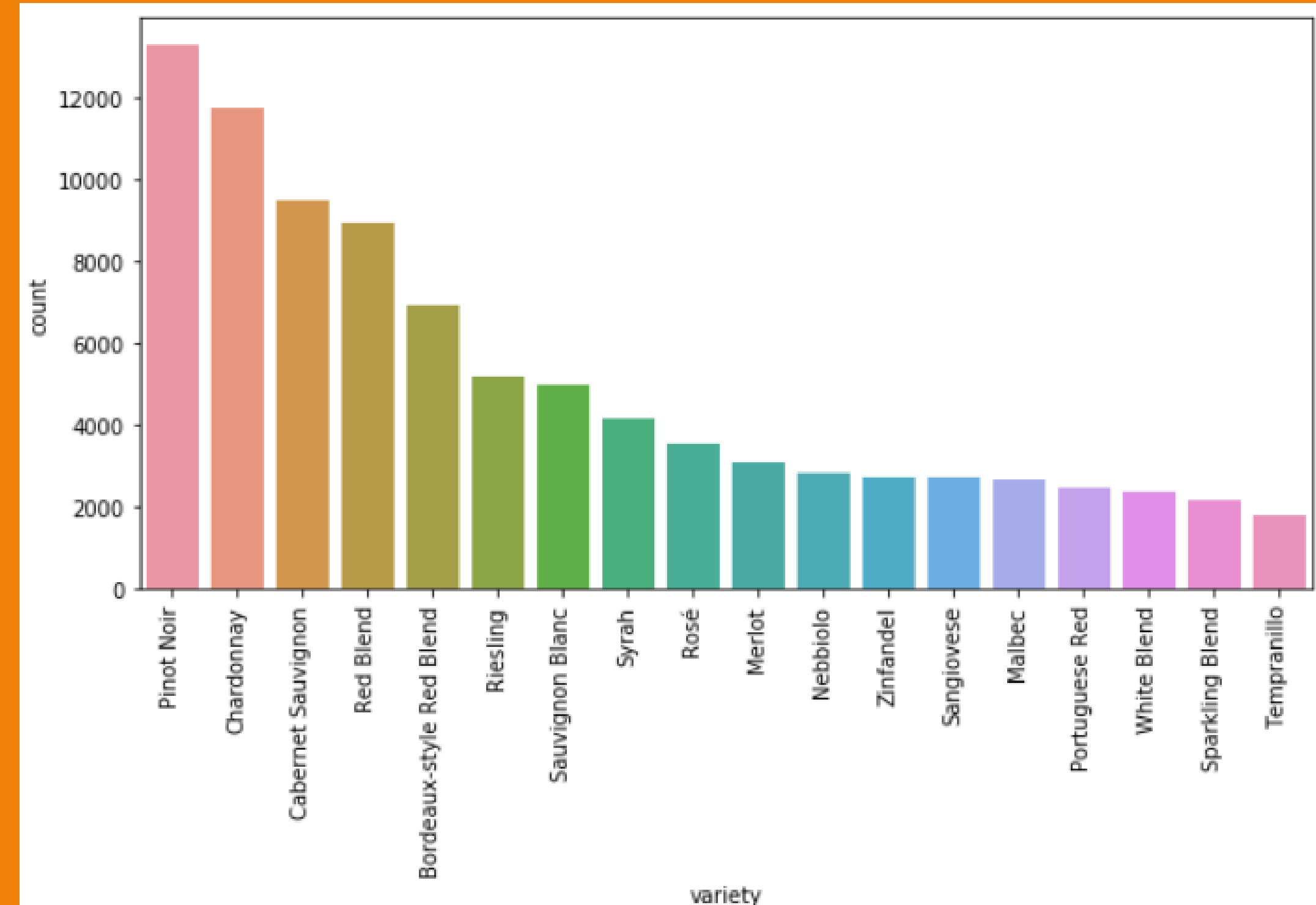
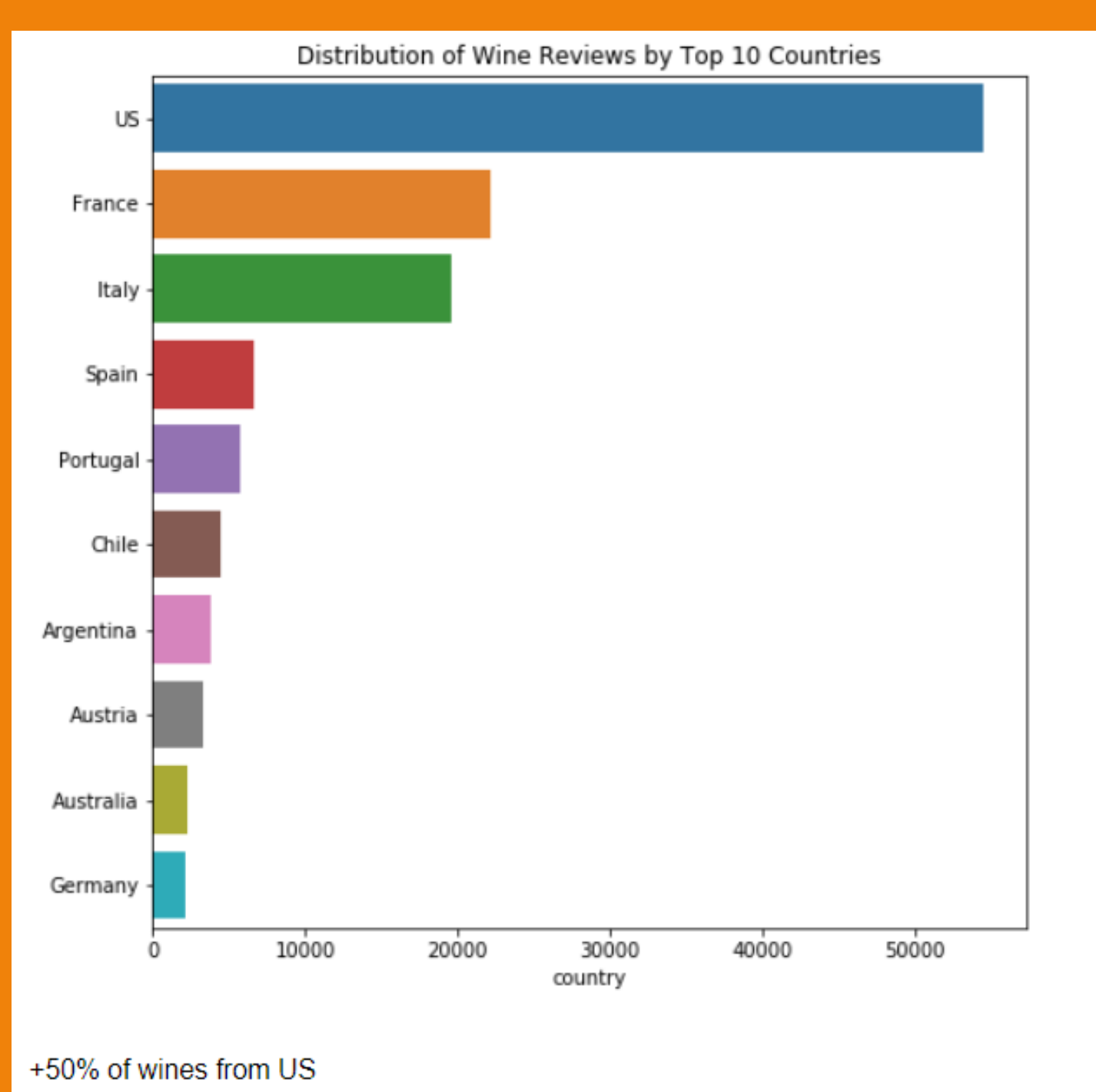
DATA CLEANING

```
df.isnull().sum()
```

```
country          63
description       0
designation      37465
points           0
price            8996
province         63
region_1         21247
region_2        79460
taster_name      26244
taster_twitter_handle  31213
title            0
variety          1
winery           0
dtype: int64
```

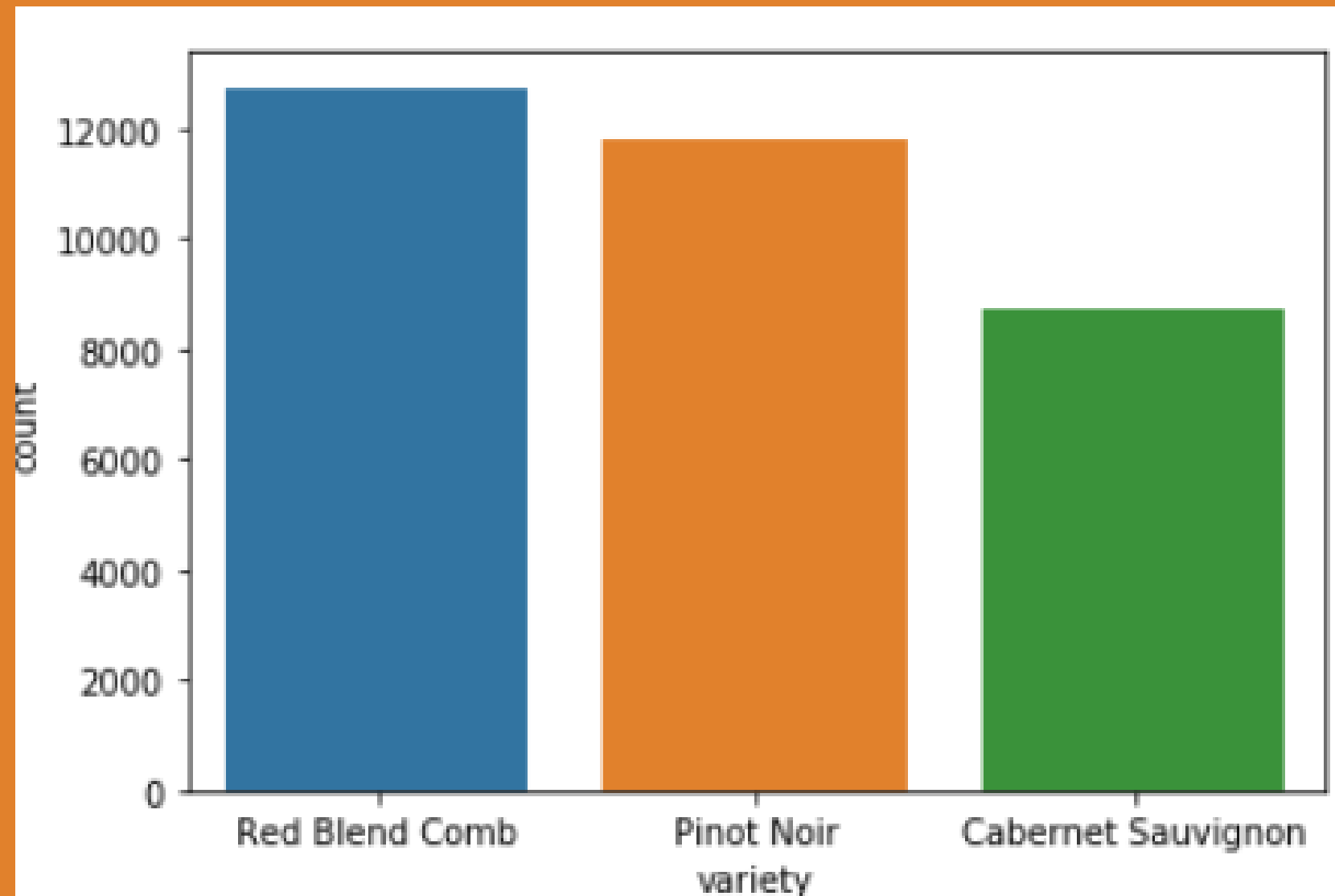
Large number of missing values in different features

VISUALIZATIONS

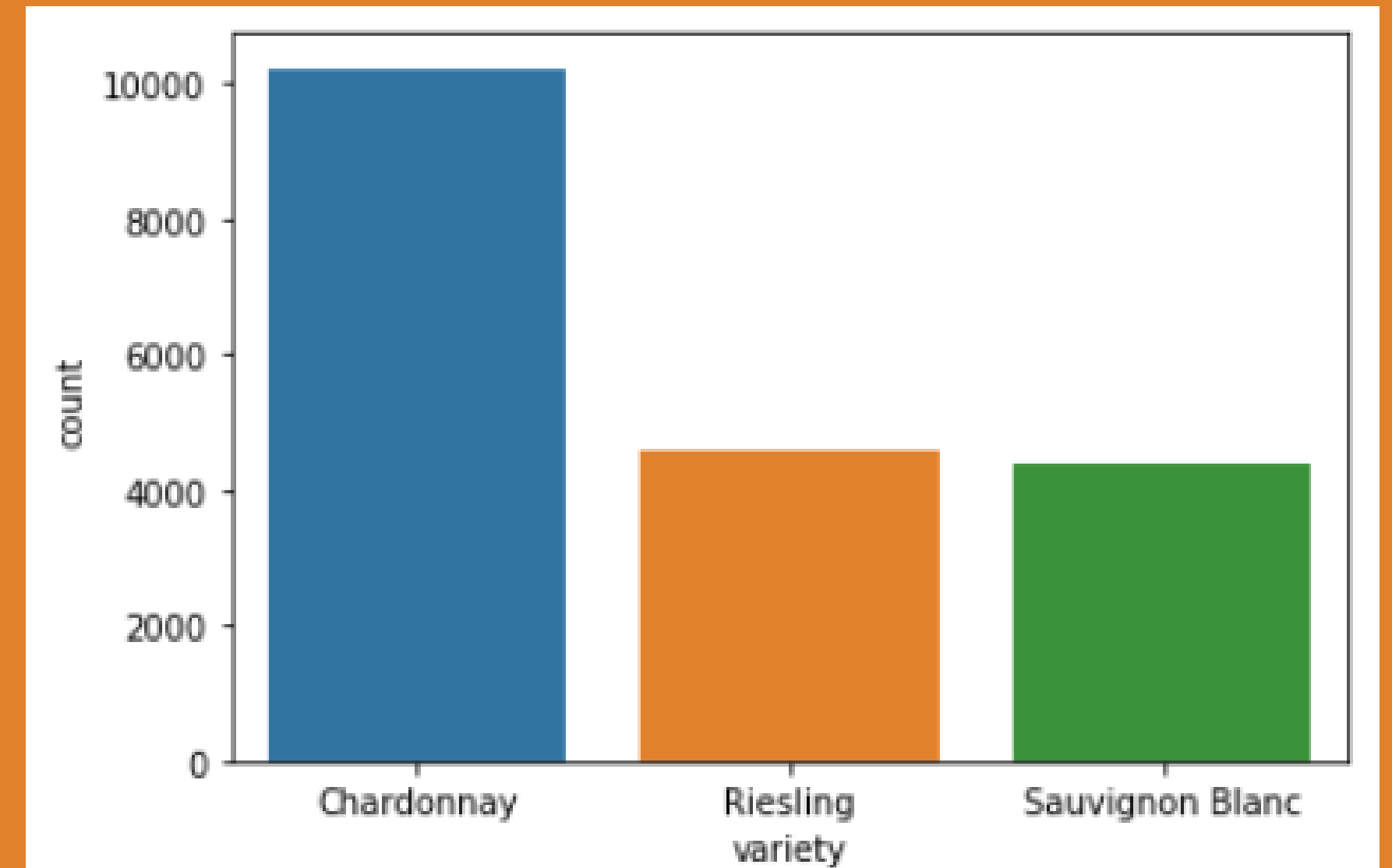


On the left top 6 grapes by the variety. On the right we combined Red Blend and Bordeaux-Style into one class because they are very similar and also to have an equal number of red and white wine types.

VISUALIZATIONS



Cabernet Sauvignon has around **4000 samples** less than Red Blend combination



Here the problem is worse, Chardonnay has **double number** of samples from other classes.

DATA PREPROCESSING

Divided the dataset into red wines and white wines



```
red_wine_df.head()
```

	country	description	points	price	province	title	variety
0	US	Much like the regular bottling from 2012, this...	87	65.0	Oregon	Sweet Cheeks 2012 Vintner's Reserve Wild Child...	Pinot Noir
1	US	Soft, supple plum envelopes an oaky structure ...	87	19.0	California	Kirkland Signature 2011 Mountain Cuvée Caberne...	Cabernet Sauvignon
2	US	Slightly reduced, this wine offers a chalky, t...	87	34.0	California	Louis M. Martini 2012 Cabernet Sauvignon (Alex...	Cabernet Sauvignon
3	US	Ripe aromas of dark berries mingle with ample ...	87	23.0	Virginia	Quiévreumont 2012 Vin de Maison Red (Virginia)	Red Blend Comb
4	US	A sleek mix of tart berry, stem and herb, alon...	87	20.0	Oregon	Acrobat 2013 Pinot Noir (Oregon)	Pinot Noir

```
white_wine_df.head()
```

	country	description	points	price	province	title	variety
0	US	Pineapple rind, lemon pith and orange blossom ...	87	13.0	Michigan	St. Julian 2013 Reserve Late Harvest Riesling ...	Riesling
1	US	Building on 150 years and six generations of w...	87	12.0	California	Mirassou 2012 Chardonnay (Central Coast)	Chardonnay
2	Germany	Zesty orange peels and apple notes abound in t...	87	24.0	Mosel	Richard Böcking 2013 Devon Riesling (Mosel)	Riesling
3	US	This shows a tart, green gooseberry flavor tha...	86	20.0	California	Envolve 2011 Sauvignon Blanc (Sonoma Valley)	Sauvignon Blanc
4	US	The clean, brisk mouthfeel gives this slightly...	86	14.0	California	Robert Hall 2011 Sauvignon Blanc (Paso Robles)	Sauvignon Blanc

Red and white wines dataframes.

PREDICTION AND OUTCOME



87%

NEURAL NETWORK

Due to high accuracy and responsiveness to data.

Links



<https://wine-recomender.herokuapp.com/>



<https://github.com/dannycyph3r/wine-recom>



CONCLUSION

In conclusion, the performance of neural networks is good. The results are better than expected using words that a user would likely input

It is not a good idea to give the user the freedom to input any text, the user would have a better recommendation if we limit its input showing a list of tastes/flavours the user can choose.



THANK YOU!