An Unambiguous Standard for the Certification and Quality Assessment in the Production of Wine while Keeping with Demand

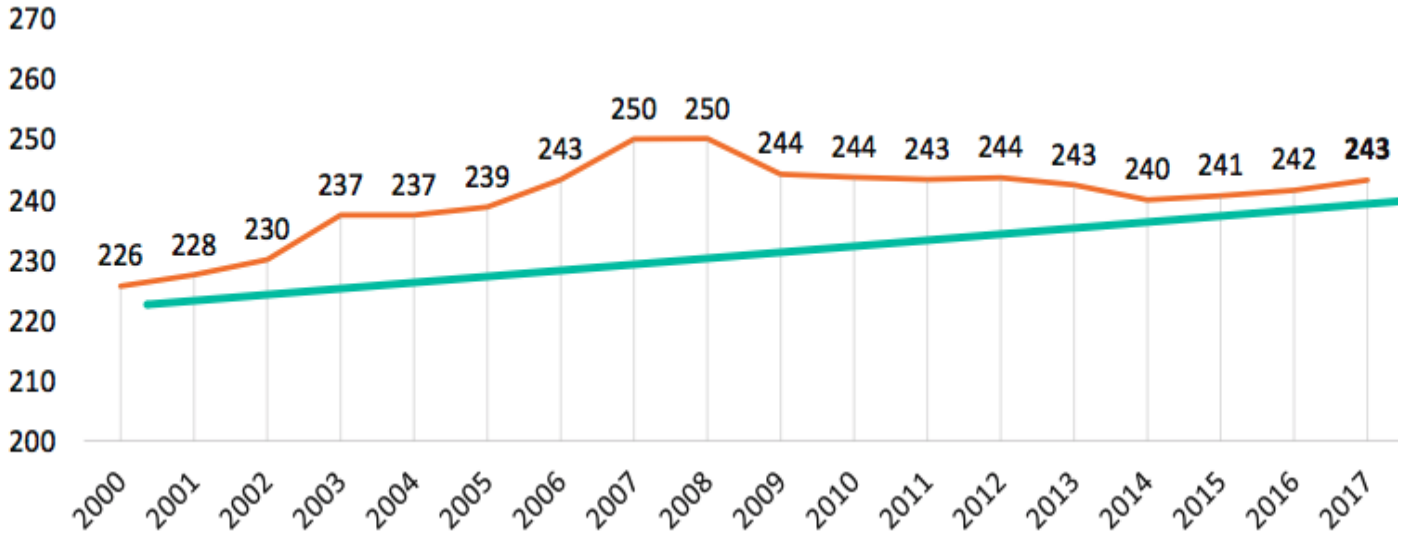
by
Tate Kennedy

September 30, 2018

The Demand for Wine

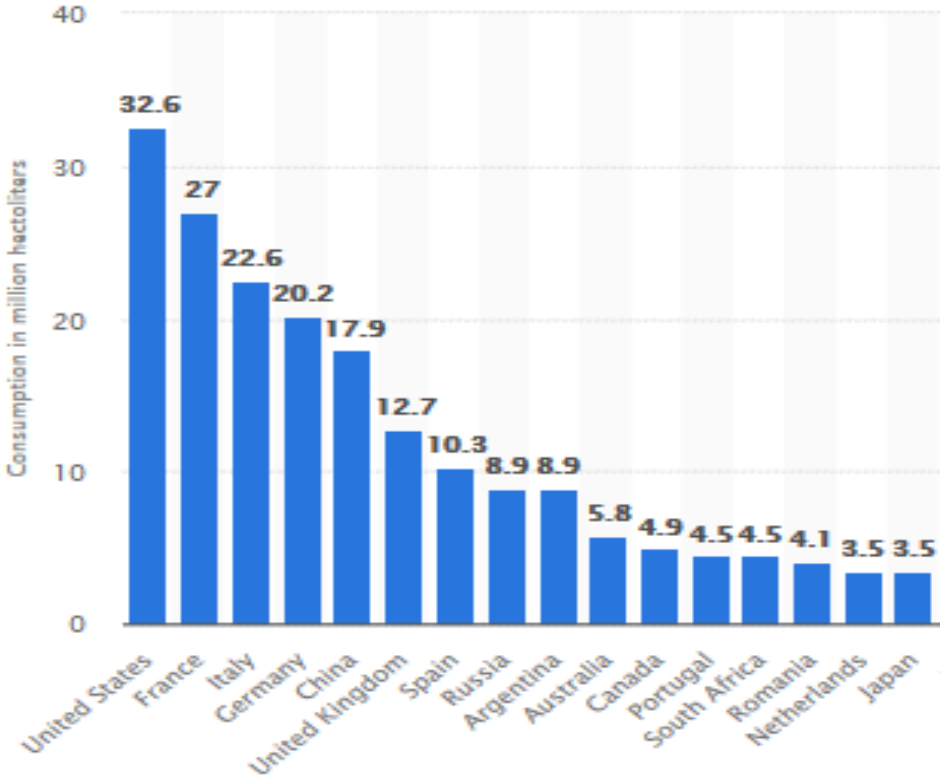
Whether it's red or white, it doesn't matter, the world can't get enough of it and we find it in almost every occasion now.

Wine Consumption Worldwide
in million hectoliters



- According to the website Statista, in 2017 wine consumption worldwide was estimated to amount to 243 million hectoliters, up from 226 million hectoliters in 2000.

Wine Consumption Worldwide in 2017
by Country (3.5 million hectoliters or greater)

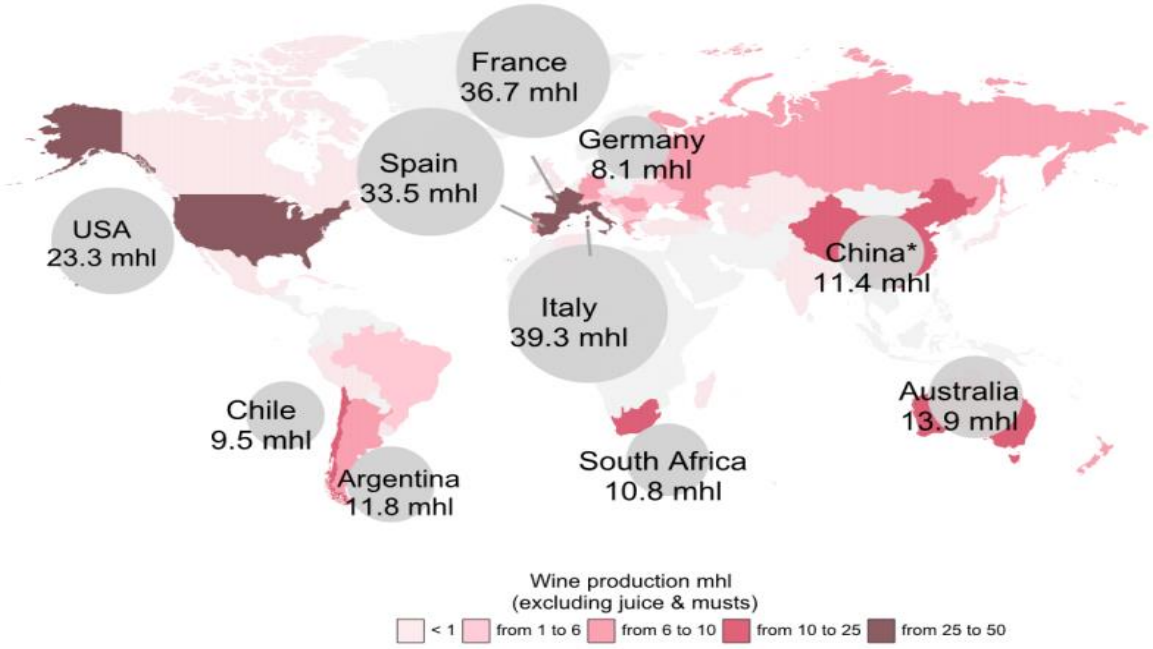


- The U.S. continues to be the largest consumer of wine globally since 2010, consuming the most wine in 2017 at 32.6 million hectoliters.

The Supply for Wine and the Challenge it Creates

After the decline following the 2008/2009 economic crisis, world wine consumption resumed a positive trend and with it, familiar concerns.

Wine Production Worldwide
in million hectoliters



- According to the website Statista, the United States became the 4th largest producer nation (following Italy, France and Spain) with over 23 million hectoliters in 2016.

- In the United States alone, wine is a $38 billion industry (GordOn, 2016).
- Wine quality is mostly assessed the more widely accepted sensory tests and by physicochemical tests.
- Sensory tests depend on human experts to sample the wine and qualify the wine according to taste.
- This subjective method is challenged in keeping up with the demands of mass production created from the increasing popularity of wine.
- The physicochemical tests can be conducted on a larger scale and have the advantage of being objective.
- So the question is, can we accurately and objectively measure the quality of wine based on physicochemical compounds while keeping with demand?

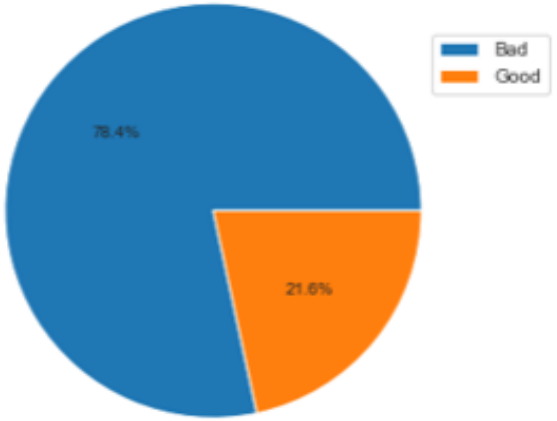
The Answer is Yes

- Over the past several weeks, I have been utilizing sensory test data and physicochemical test data that were collected for red and white variants of the Portugese “Vindho Verde” wine.
    - Datasets are available from the UCI Machine Learning repository https://archive.ics.uci.edu/ml/datasets/wine+quality. (P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009).
- The aim has been to explore the relationship between the two types of test results and whether a machine learning approach/product could predict the quality in wine in an objective, efficient and large scale cost effective manner.
- Thus, solving for the subjective, cumbersome and costly process of certifying and assessing the quality of wine in the industry’s current production model.
- Within the supervised learning branch of machine learning, I was able to use classification models to categorize the outcome of interest (quality) between discrete values such as good (quality score of 7 or higher) and bad (everything else).
- The Random Forest Classifier produced the highest accuracy rate amongst both the red wine variants (89%) and the white wine variants (87%), beating out several other candidates as the best performer.
- The model successfully interprets the physicochemical compounds found in each variant of red and white wines and predicts the quality of a particular variety.

Modeling Approach and Evaluation

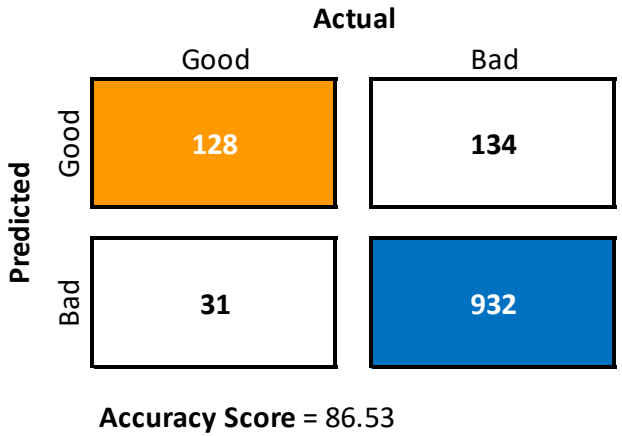
- In supervised learning, models will be trained to generate a certain outcome or dependent variable, something we observed previously (the quality of wine via sensory tests) and we want to use new inputs (physicochemical compounds/tests) to generate a prediction.
- What kind of variable the outcome of interest (quality) is informs us what kind of model to build. Since quality is a categorical variable (good or bad, a binary outcome), a classification model was built.
- I will use the example from the white wine variants to illustrate the results achieved using a Random Forest Classifier:

White Wine Quality



- The white wine dataset contained 3,838 variants ranked as Bad and 1,060 ranked as Good, in total.

Confusion Matrix



Accuracy Score = 86.53

- 1,225 samples taken as a test set.
- This matrix shows the count of each permutation of target and prediction so one can see where the model can be improved.
- The model accurately predicts quality on 1,060 samples, or at a rate of 87%.
- The model leaves room for improvement, with quality predicted incorrectly on 165 samples in total (31 false negatives and 134 false positives). Note, results were slightly better in the red wine variants.

The Value in this Approach

My product provides solutions that address those concerns well documented here but will also provide opportunities for producers to bridge gaps with consumer to better understand and deliver on their preferences:

- Within the wine industry as a whole, certification and quality assessment are those critical concerns already noted. My product provides an unambiguous standard that is needed to regulate the market with greater efficiency and lower costs.

- The model will shed light on how physicochemical compounds affect the sensory preferences of consumers. This is an area today, where that relationship is not clearly understood within the industry.

- The understanding of human preferences in sensory testing is subjective and is not enough to ensure quality wine is produced at a larger scale, in keeping with demand.

- The model will also let users identify the most important features or compounds out of the those available for producing quality wine. It offers an experimental method for detecting variable interactions.

- There is an entire consumer application to this approach/product as well I haven't discussed. Imagine a web or mobile app that grows and learns with users by accessing these same physicochemical tests to provide recommendations based on objective results rather than the subjective nature of today's apps. That same information could be passed through to companies opening several new doors!