

## Problem statement

Given the price, year made, country, region, and variety of a bottle of wine, predict the points of this bottle.

## Instructions

1. You may need to install a few Python libraries, including sklearn, scipy, and matplotlib.
2. DIC\_Phase1 and DIC\_Phase2\_AND\_Phase3 are Jupyter files. First, run DIC\_Phase1.ipynb by clicking on “Run All” in your code editor (or however you run Jupyter notebooks on your system). Then, also run DIC\_Phase2\_AND\_Phase3.ipynb **except the last code block**. Important: In DIC\_Phase2\_AND\_Phase3.ipynb, run everything but the last code block because that last code block is how you run the GUI.
3. When you’re ready, you may run the last code block from DIC\_Phase2\_AND\_Phase3.ipynb and that should spawn a GUI.

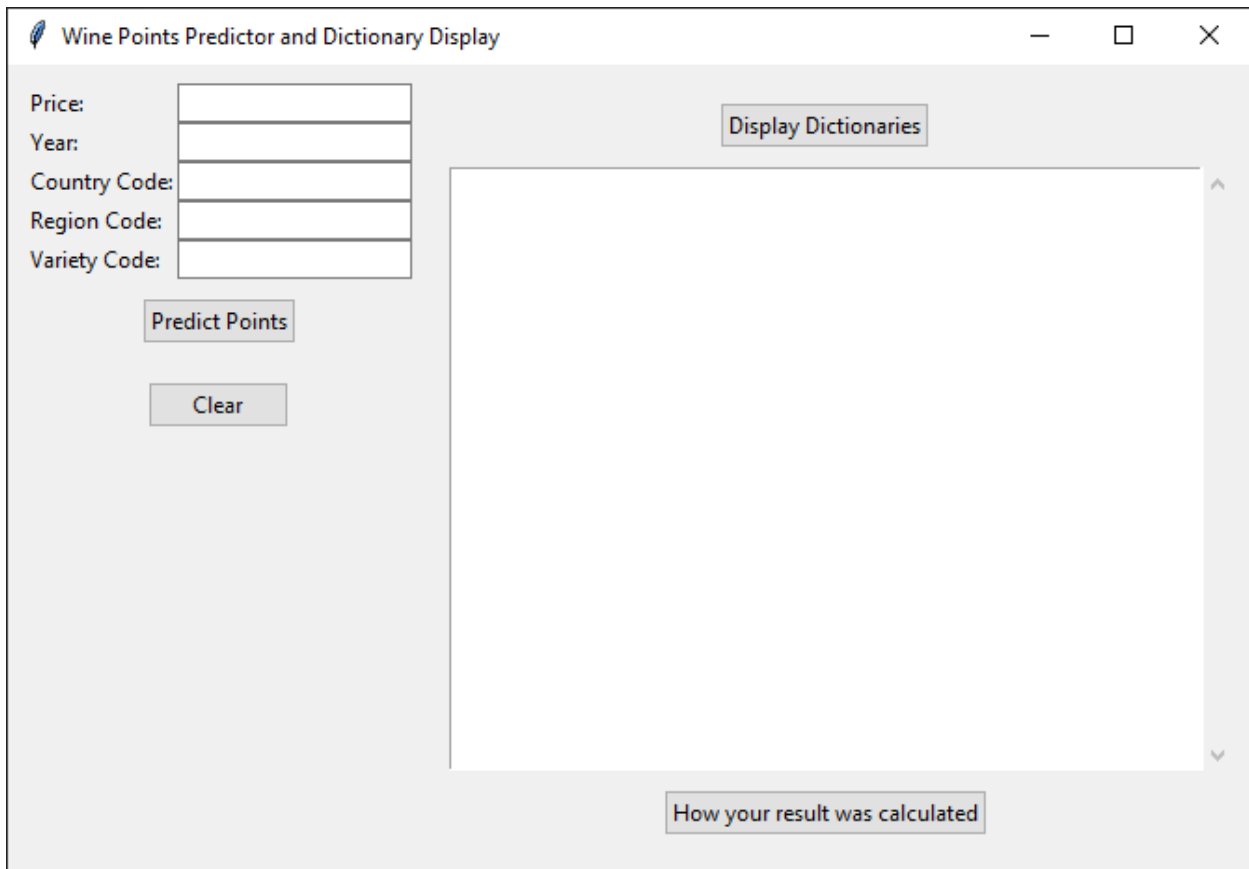


Figure 1

Wine Points Predictor and Dictionary Display

Price:

Year:

Country Code:

Region Code:

Variety Code:

Predict Points

Clear

Display Dictionaries

Country Codes:

- US: 0
- Spain: 1
- Italy: 2
- France: 3
- Argentina: 4
- Australia: 5
- Canada: 6

Region Codes:

- Willamette Valley: 0
- Lake Michigan Shore: 1
- Navarra: 2
- Vittoria: 3
- Alsace: 4
- Napa Valley: 5
- Alexander Valley: 6
- Central Coast: 7
- Cafayate: 8
- Mendoza: 9

How your result was calculated

Figure 2

Wine Points Predictor and Dictionary Display

Price: 14

Year: 2014

Country Code: 0

Region Code: 0

Variety Code: 0

Predict Points

Clear

Predicted points: 87.54

Display Dictionaries

Country Codes:

- US: 0
- Spain: 1
- Italy: 2
- France: 3
- Argentina: 4
- Australia: 5
- Canada: 6

Region Codes:

- Willamette Valley: 0
- Lake Michigan Shore: 1
- Navarra: 2
- Vittoria: 3
- Alsace: 4
- Napa Valley: 5
- Alexander Valley: 6
- Central Coast: 7
- Cafayate: 8
- Mendoza: 9

How your result was calculated

Figure 3

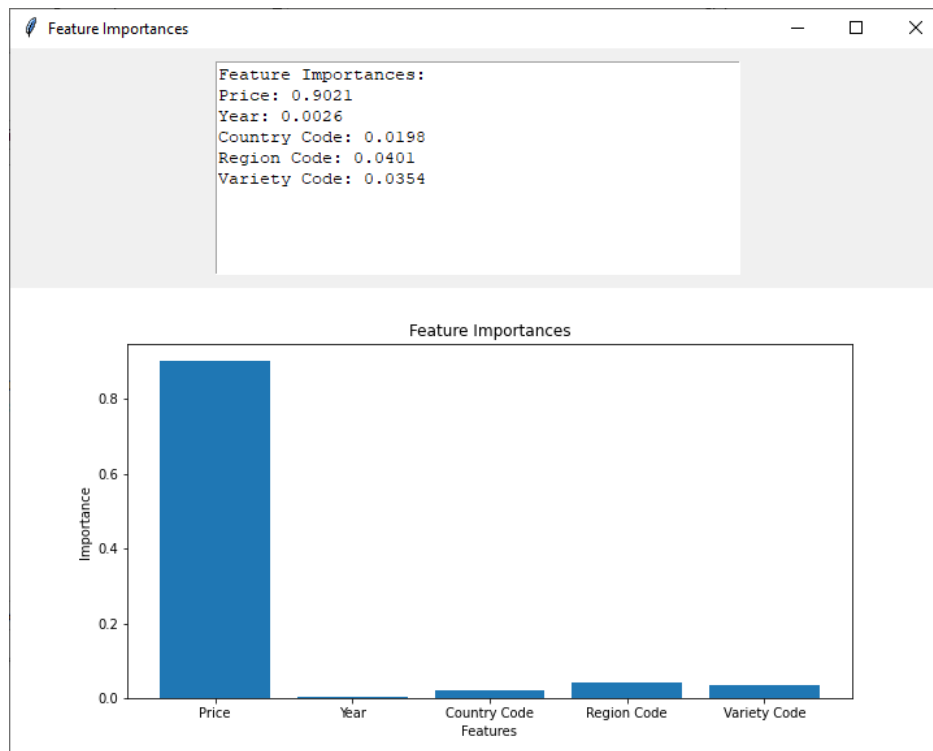


Figure 4

- At the first boot, figure 1 is what you'll see. Click on "Display dictionary" on the top right, and you'll see something like figure2. This dictionary holds the country, region, and variety codes. Scroll through this list, find the corresponding code to your query, and use that code in the input fields on the left. Note that "price" and "year" are up to the user. Once you've filled out the input fields, click on "Predict Points" and receive the points predicted by the model as in Figure 3. Hit clear to clear the input fields or modify the input fields to start a new query. Hit "How your result was calculated" to see Figure 4, which tells you how the predicted points score from before was calculated.

## Notes

I ended up using my last model, the decision tree model. Even though the decision tree regressor didn't yield a great result, at a measly 37.5% precision, it was one of the best scores compared to my other models. The poor accuracy was probably because of the large variation that can be seen in from the graph at "Basic Linear Regression 1/5". After brute forcing "max\_depth" variable under the DecisionTreeRegressor initialization from 5-20, I tuned this depth to 8, since it was the number that maximized my score. Besides that, I followed the convention of training size at 75% of my data set, and 25% for tests.

## Recommendations

Let's assume the higher the points, the better a bottle of wine is, and conversely the lower the points, the worse a bottle of wine is. My data product is a way for wholesalers and consumers of wine to predict the "goodness" of a bottle of wine based on certain objective metrics. On testing of a few inputs, the predictor performed within 3 points of the actual points for input, which is surprisingly good considering

how badly it performed during regression. What we can glean from this fact is that the “goodness” of wine is subjective, that is, an expensive bottle of wine may or may not have higher points than another with a lower price, as proven from the variation when I plotted price to points. To extend this project, perhaps an avenue would be to add more data into the modelling – if you can’t do quality, you can try quantity. By swapping points and price (that is, predict price using points, country, region, variety, and year), wholesalers and consumers could also get a “recommended” price for the bottle they’re eyeing.