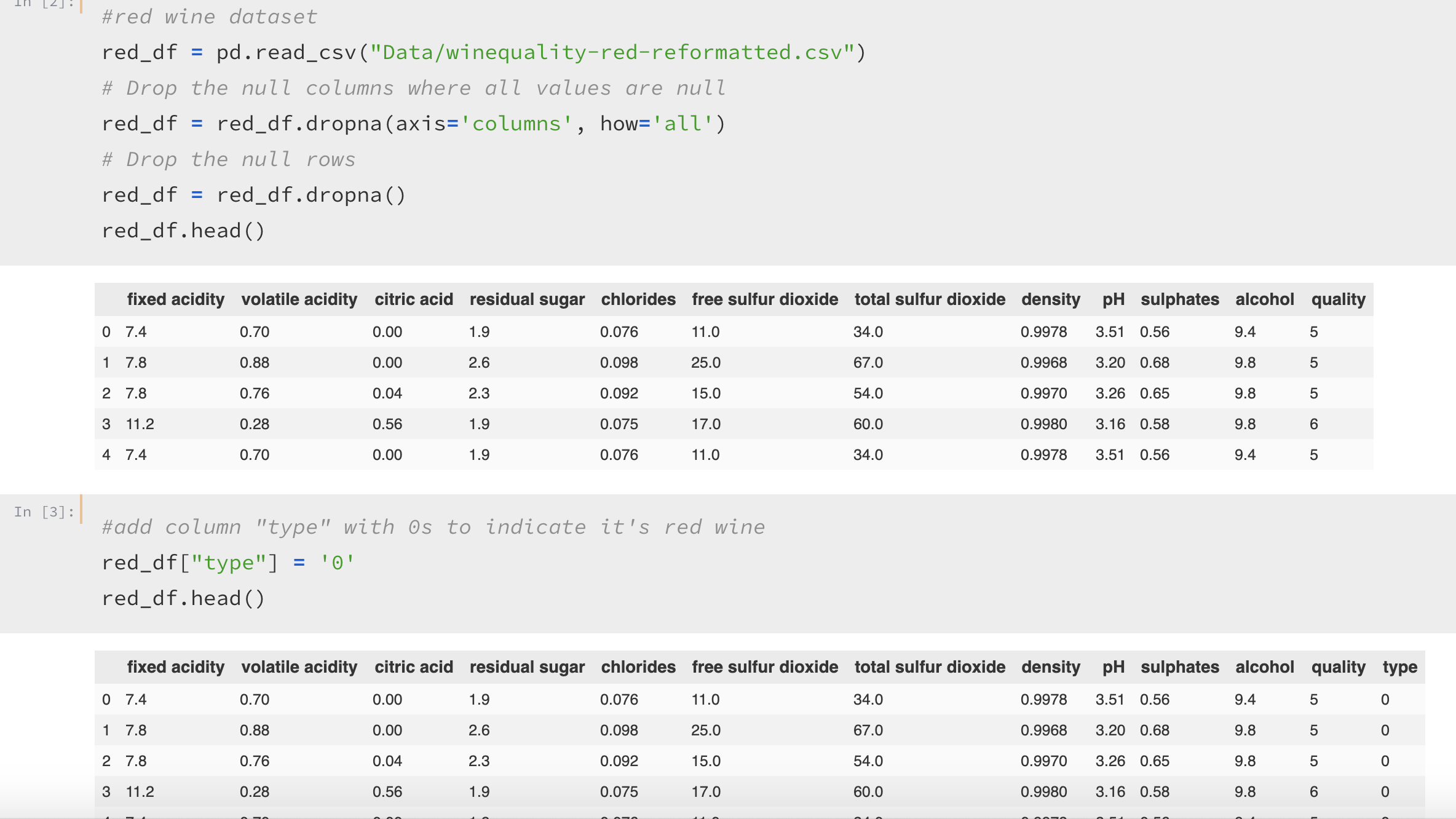
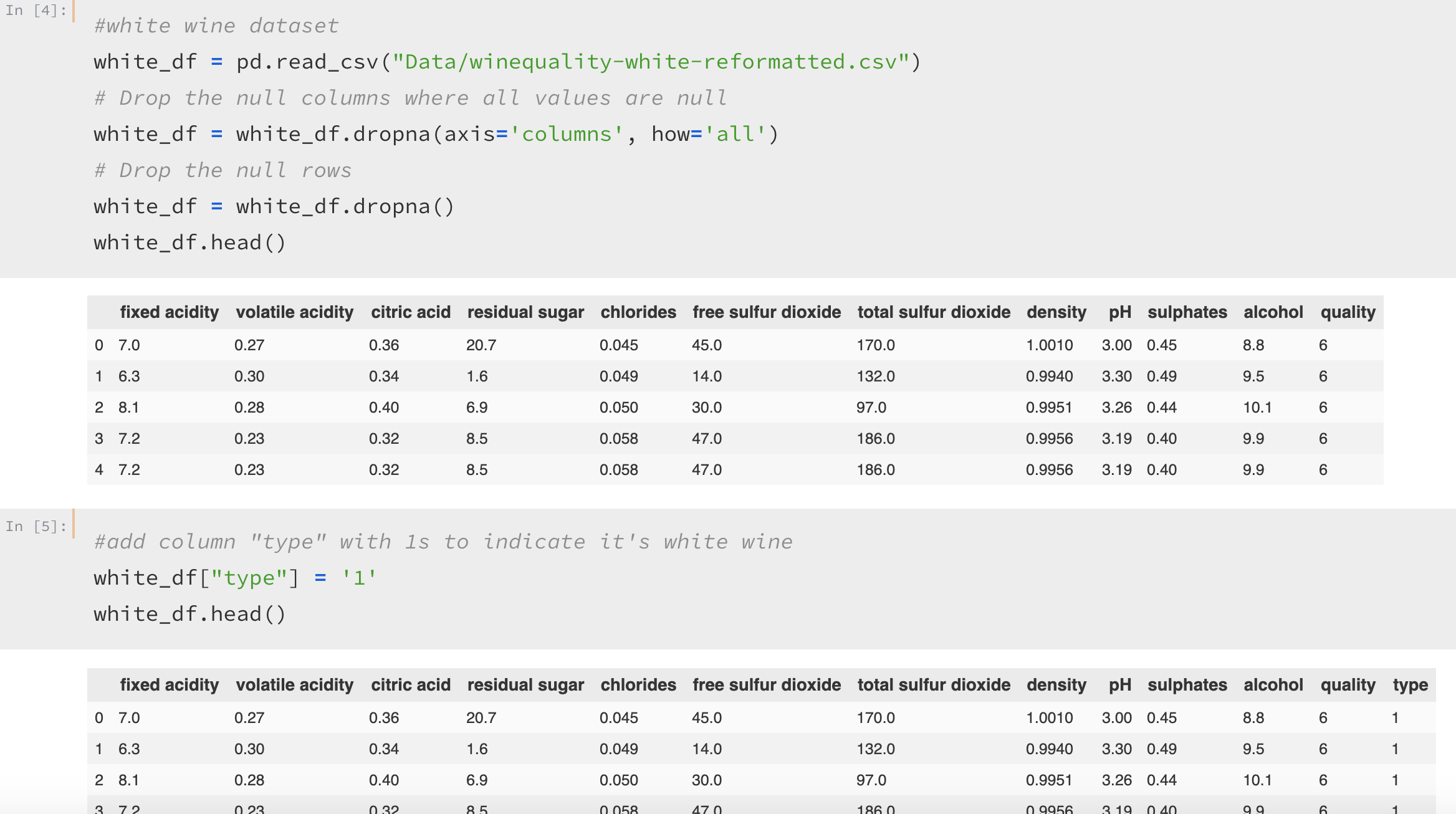
RED VS. WHITE WINE

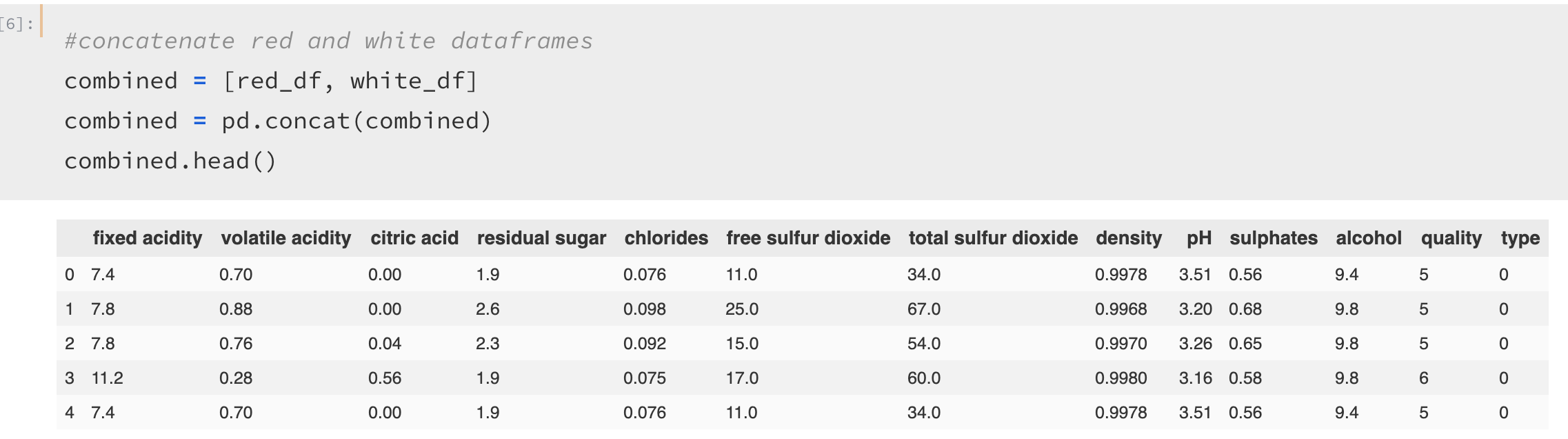
**Objective**: Building a classification model that would predict whether a wine is red or white based on physicochemical properties: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol.

**Combining the red and white dataset**:

* Adding a new column named “type” and add “0” as values to indicate the wine is red.

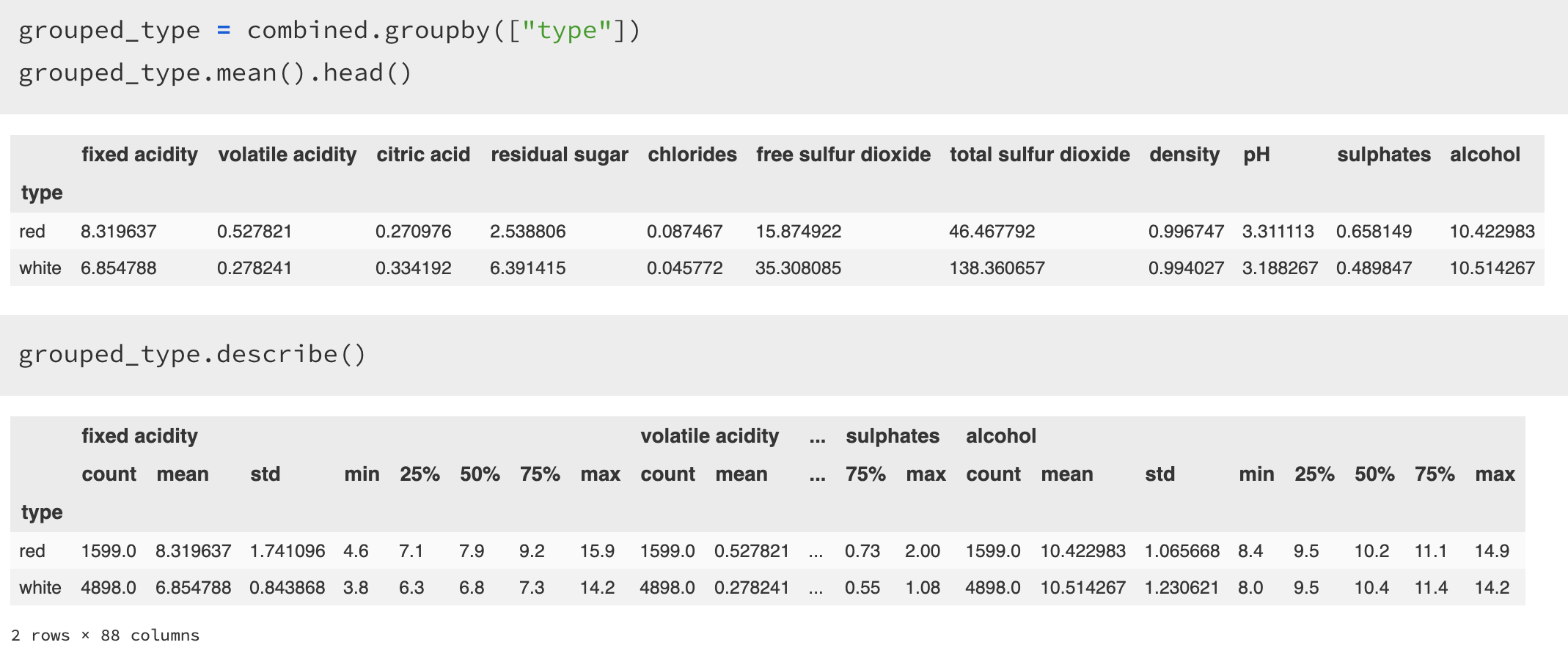


* Same goes for white, but add “1” instead.

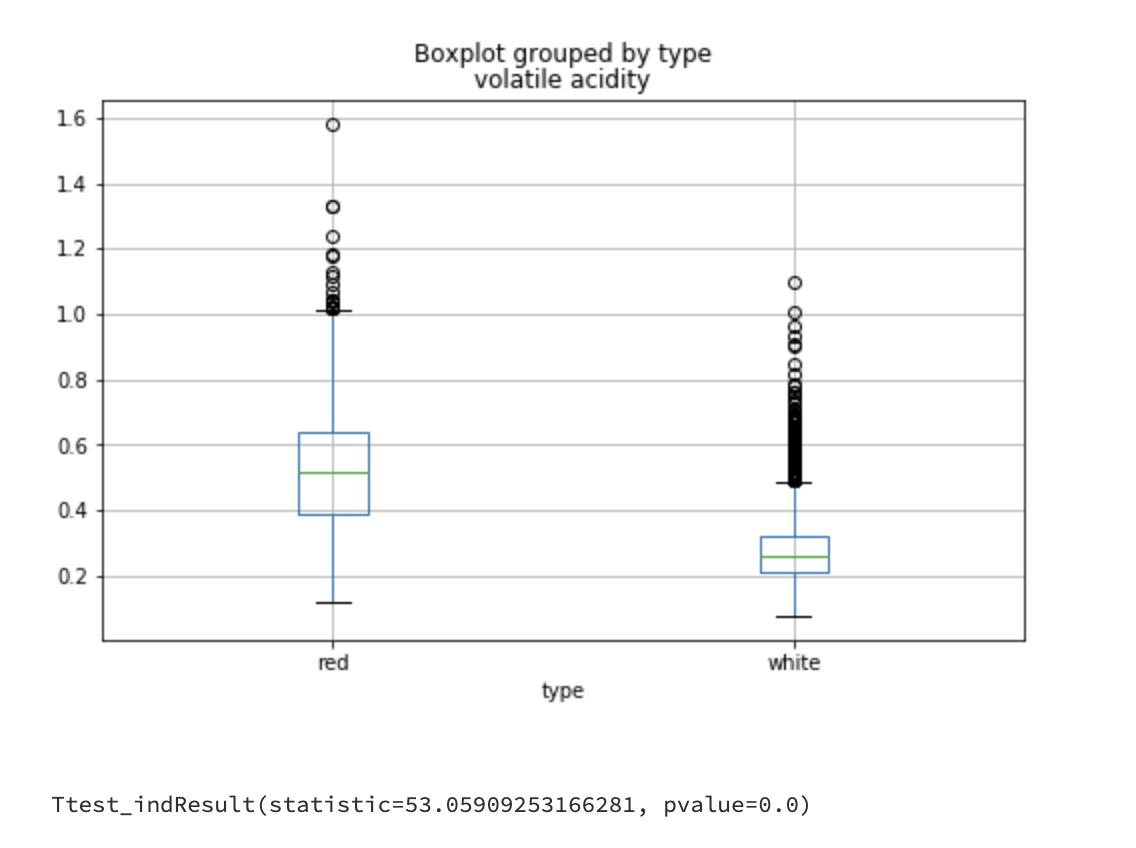
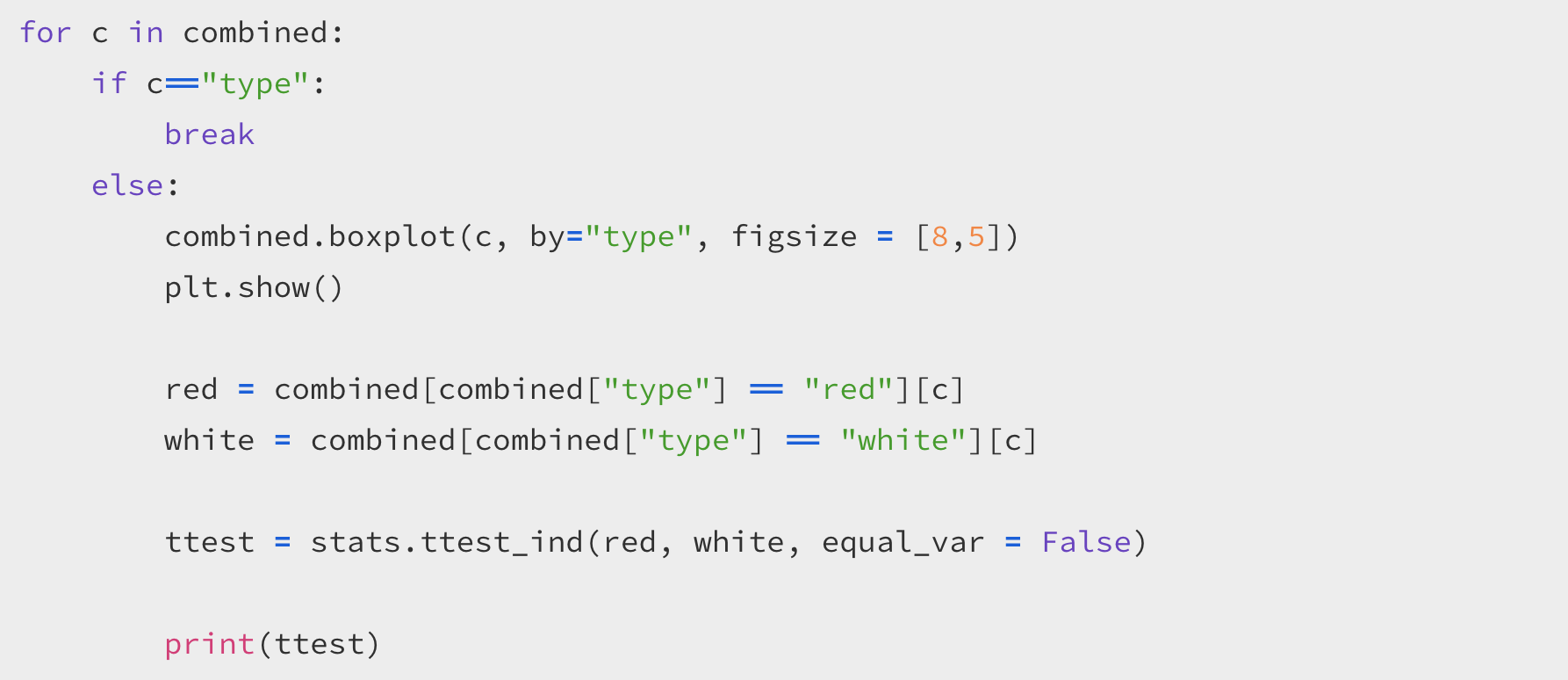


* Concatenating red and white dataset for our analysis

**Exploring the combined dataset**:

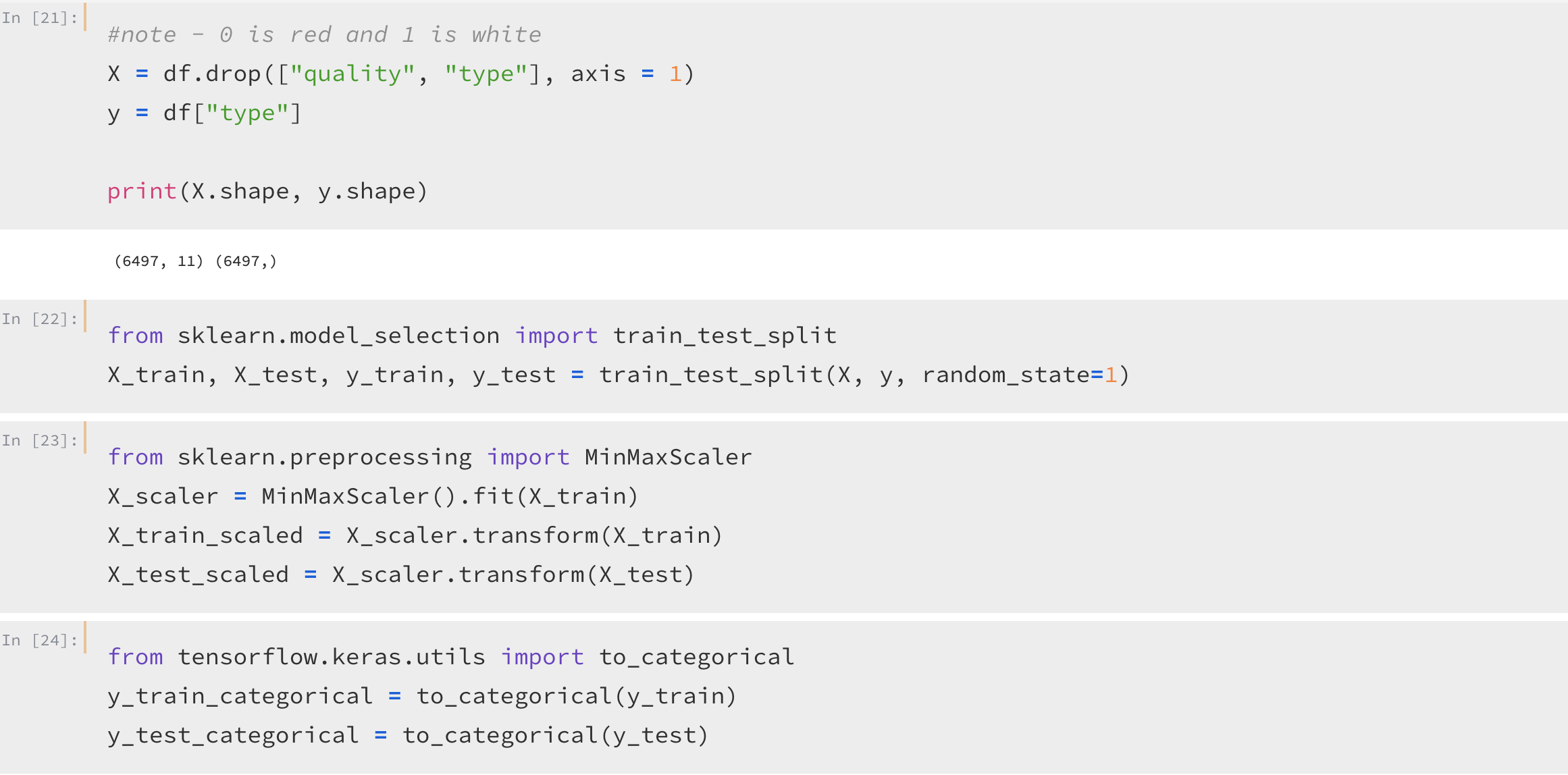


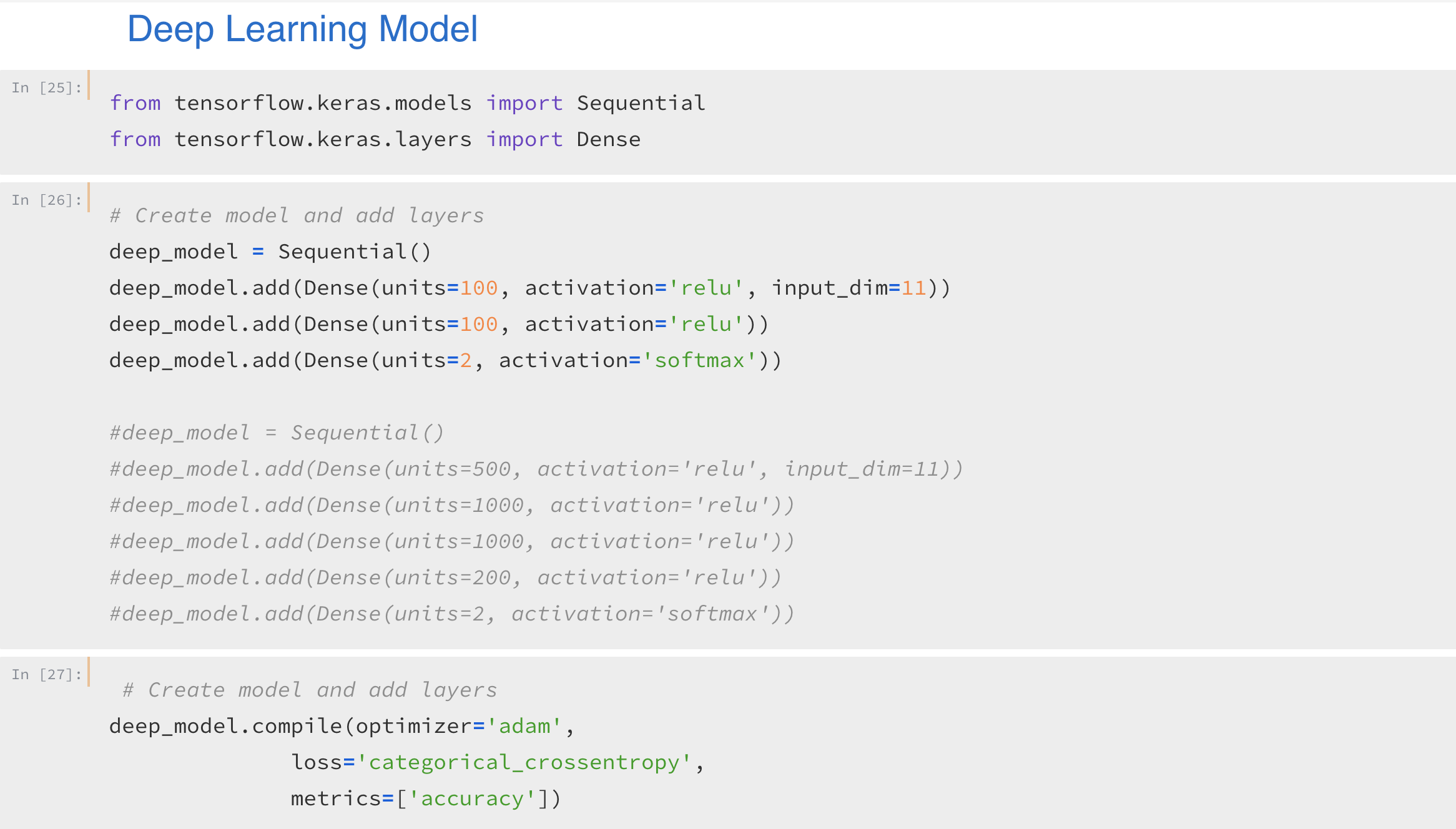
* Grouping the dataset by “type” to further examine the other columns (physiochemical properties) and see if we see any difference between the two groups.
* Mainly, looked into their averages for each column.

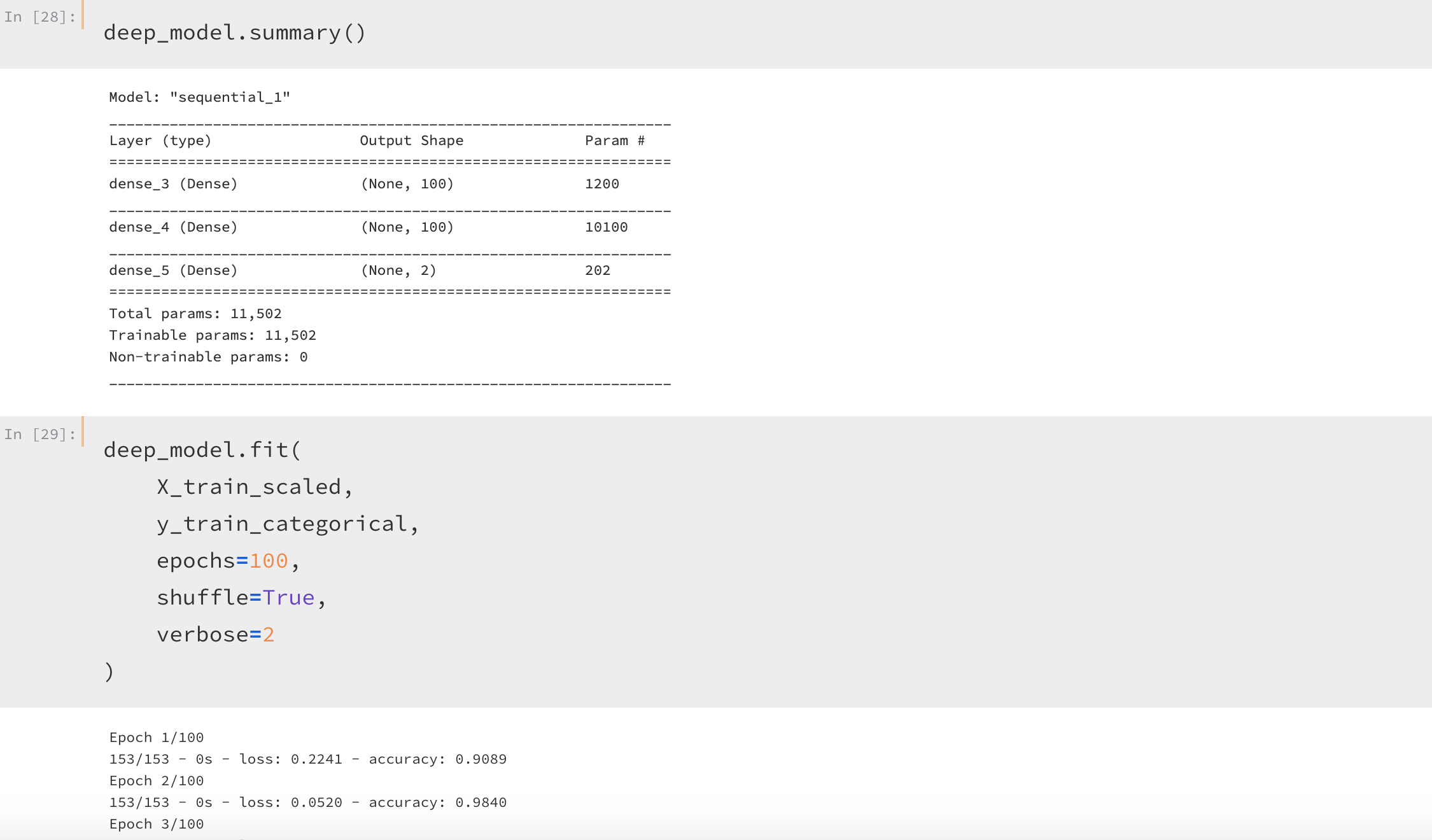


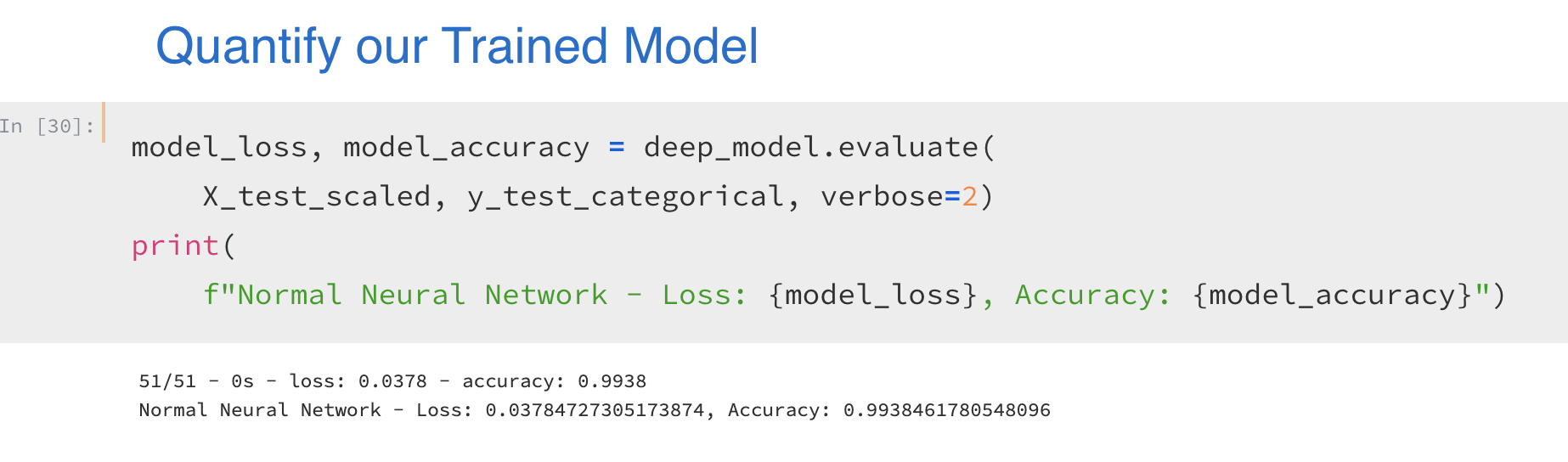
* Conducted a ttest to determine for each physicochemical property if there is a statistical difference red and white wine.
* The box plot above just shows for volatile acidity, but all the other physicochemical properties were also significantly significant (p < 0.05) as well.

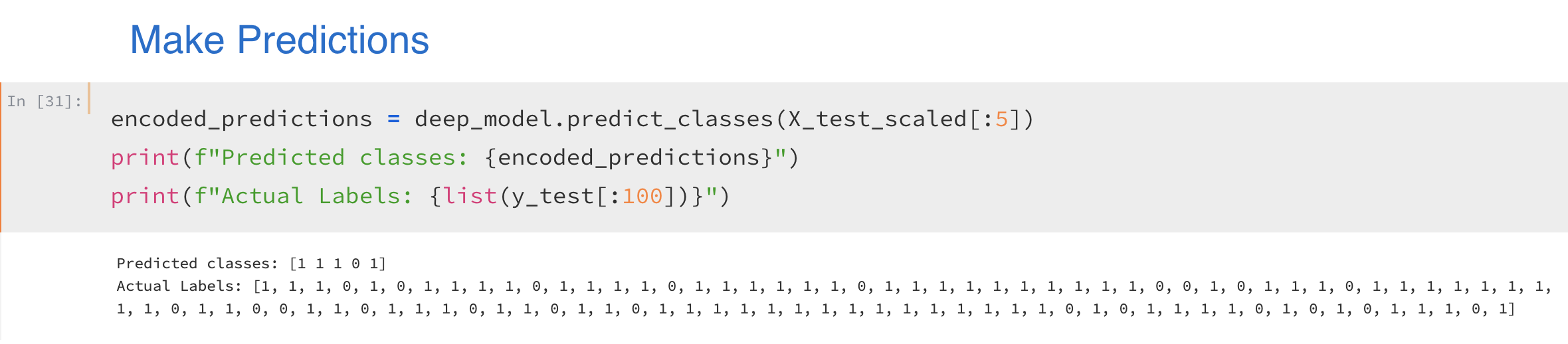
**Building Deep Learning Model**:



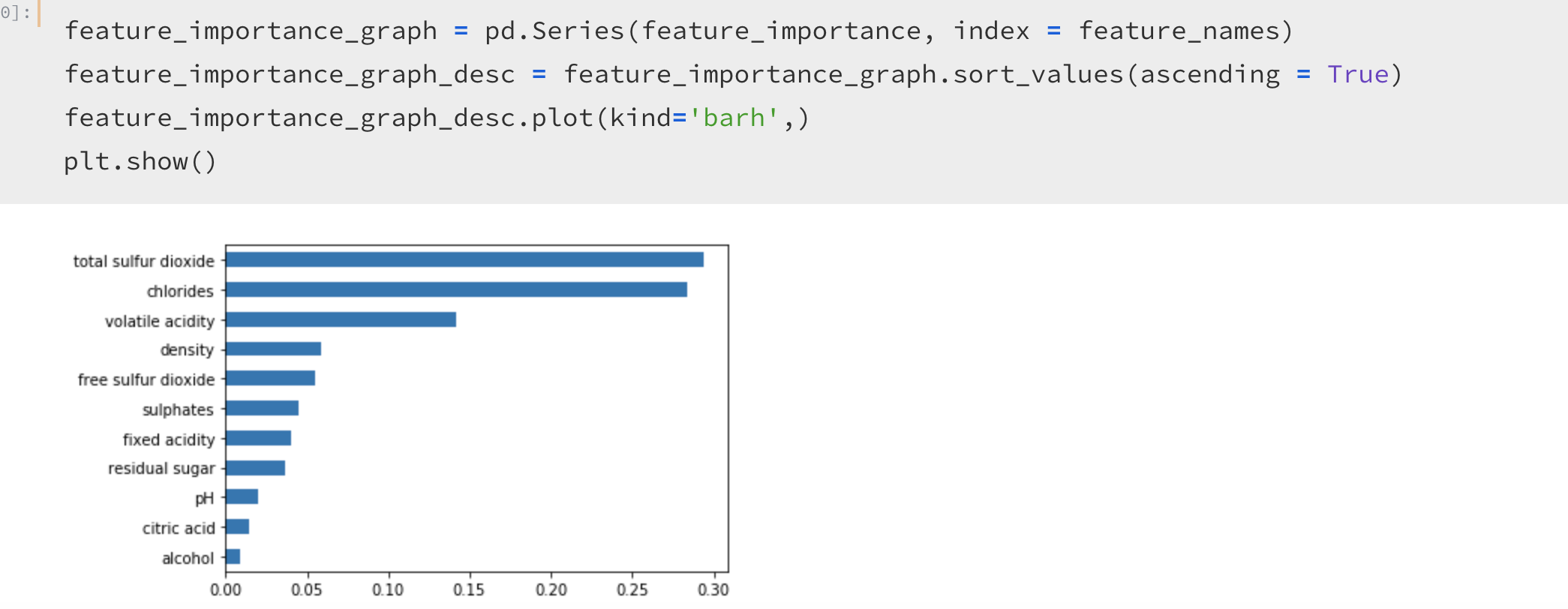
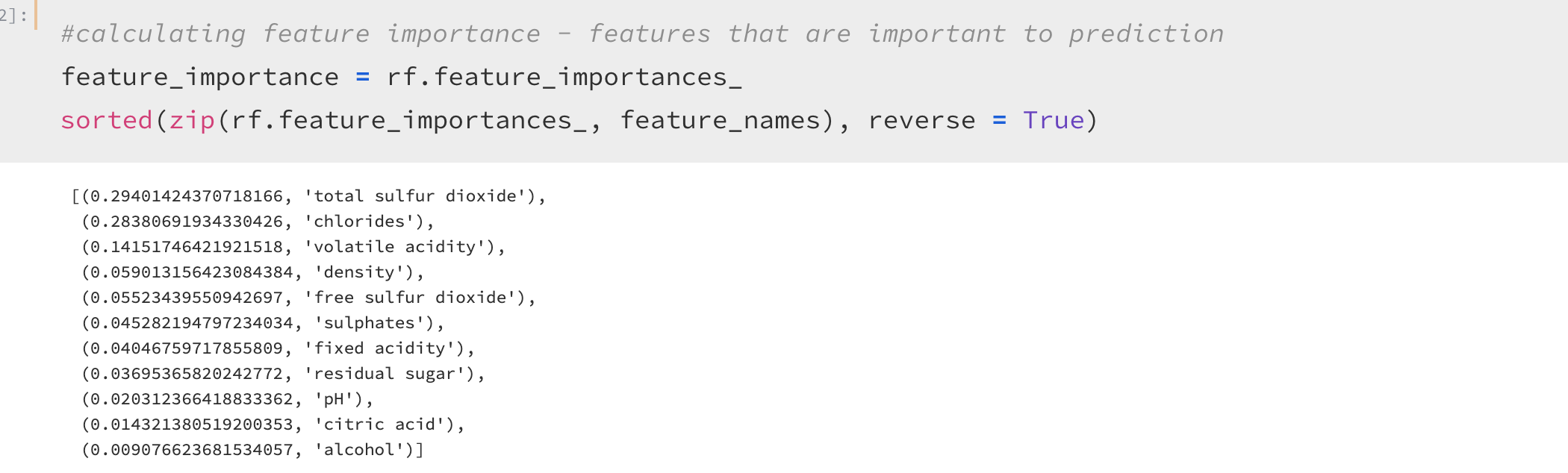








* Created a deep learning model – 11 inputs, 2 layers of 100 hidden nodes, and 2 outputs (red: 0, white: 1).
* Achieved a model with really high accuracy (close to 1) – no need for extra training.
* Save the model as a “h5” file that we load into Flask.

**Feature Importance**:

* Using Random Forest to calculate feature importance to determine which features are important in predicting a wine being red or white.