

**London College of Professional Studies**

***Assignment Cover Page***

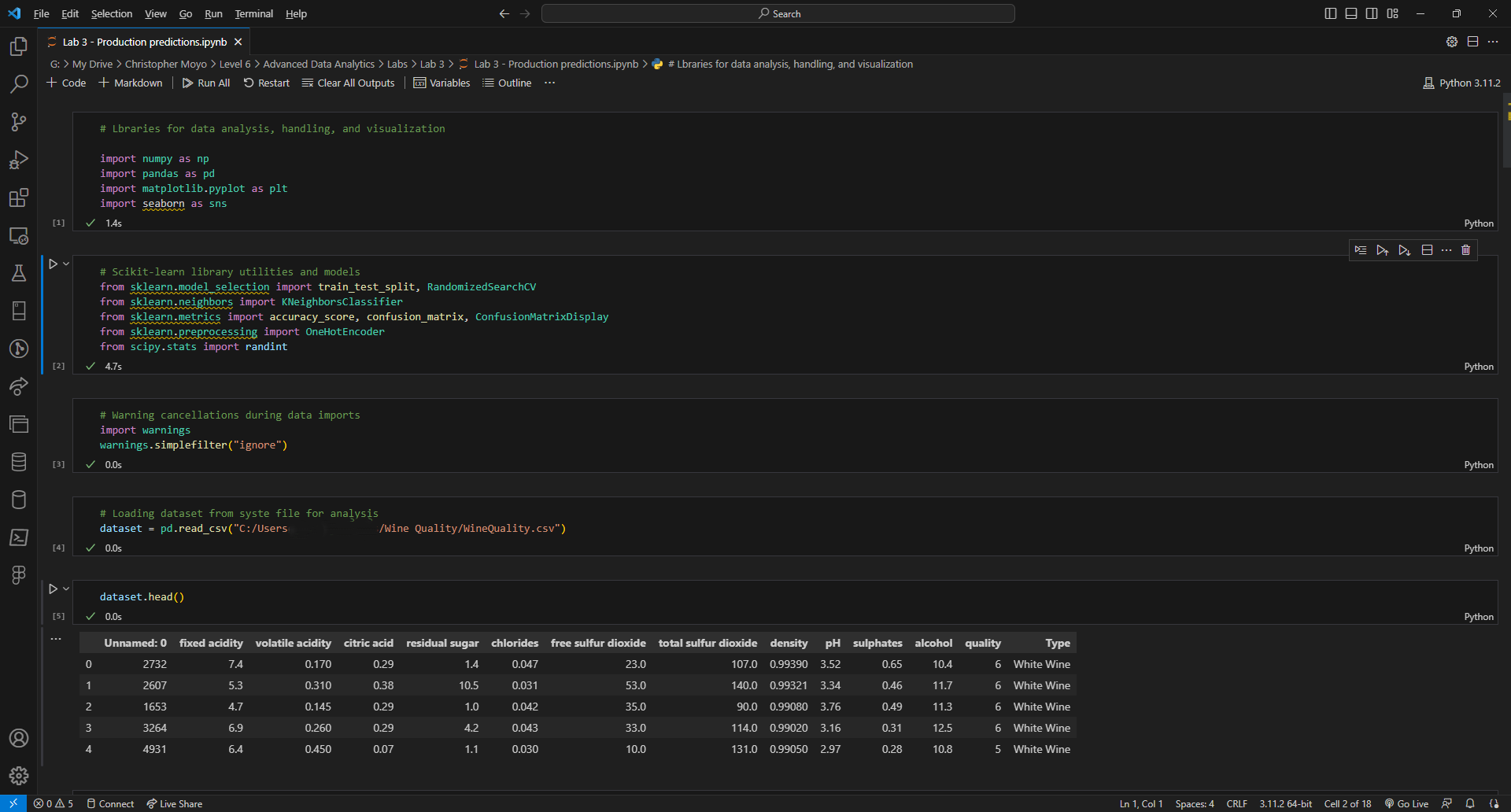
|  |  |
| --- | --- |
| **Student name** | **Christopher Moyo** |
| **Student ID** | **14034** |
| **Submission date** | **28/08/2023** |
| **Qualification title** | **OTHM Level 6 Diploma in Information and Technology** |
| **Qualification code** | **603/3789/8** |
| **Unit title** | **Advanced Data Analytics** |
| **Unit Reference Number** | **Y/617/3035** |

 I declare that the attached work is entirely my own and that all sources have been acknowledged ☒

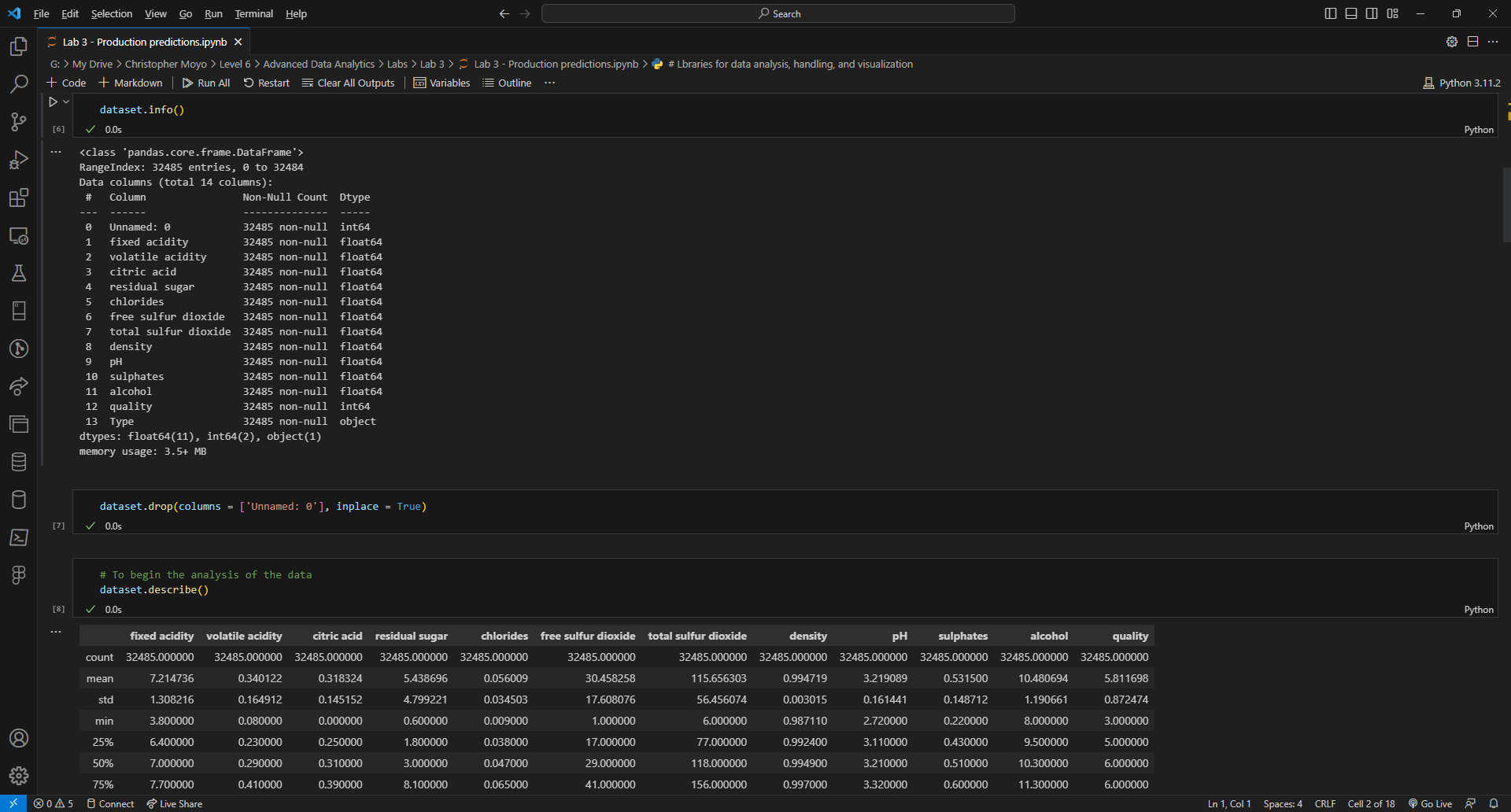
Our final predictive analysis was performed on the production of wine. The focus of this analysis was the quality of the produced wine by type before its distribution to stores or customers. The tools used were Microsoft Visual Studio Code, Jupyter Notebook, and Python programming language. Below is the demonstration.

The used dataset is borrowed from Cyber Cop on the Kaggle data analytics platform.

1. Adding required libraries and getting the relevant dataset file read into the notebook for analysing, as well as checking the brief of the dataset.

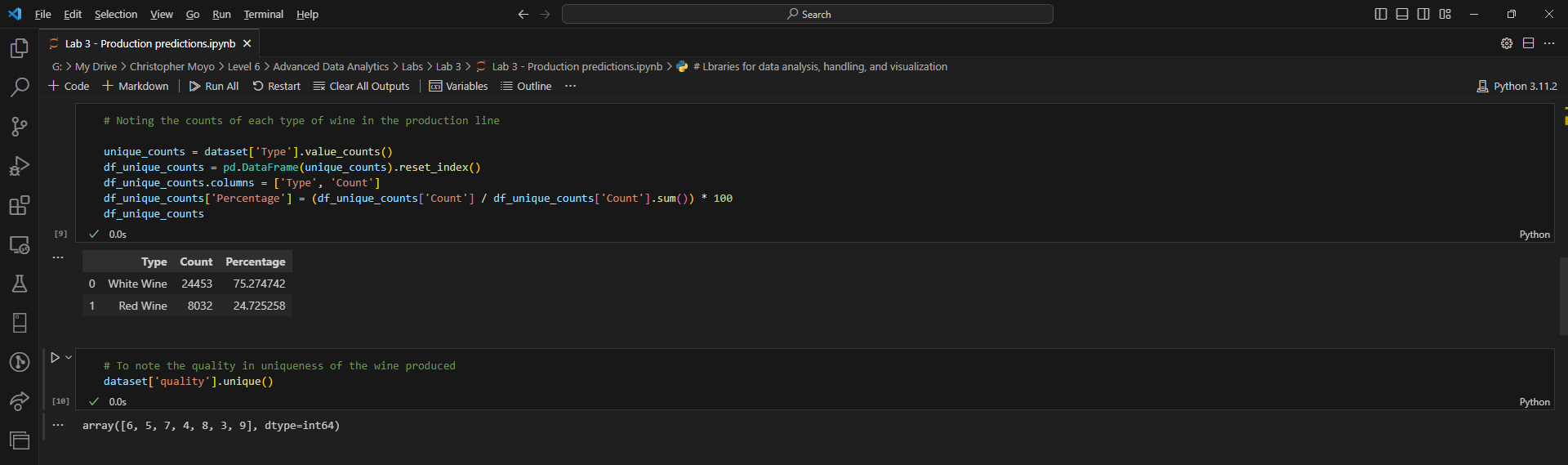


* We begin by importing the relevant libraries we are to use for analysis in cells 1, 2, and 3, and then prompt the pandas library to read our dataset in cell 4.
* Cell 5 assist us in knowing the first 5 members of the dataset and the description of the dataset we are using.



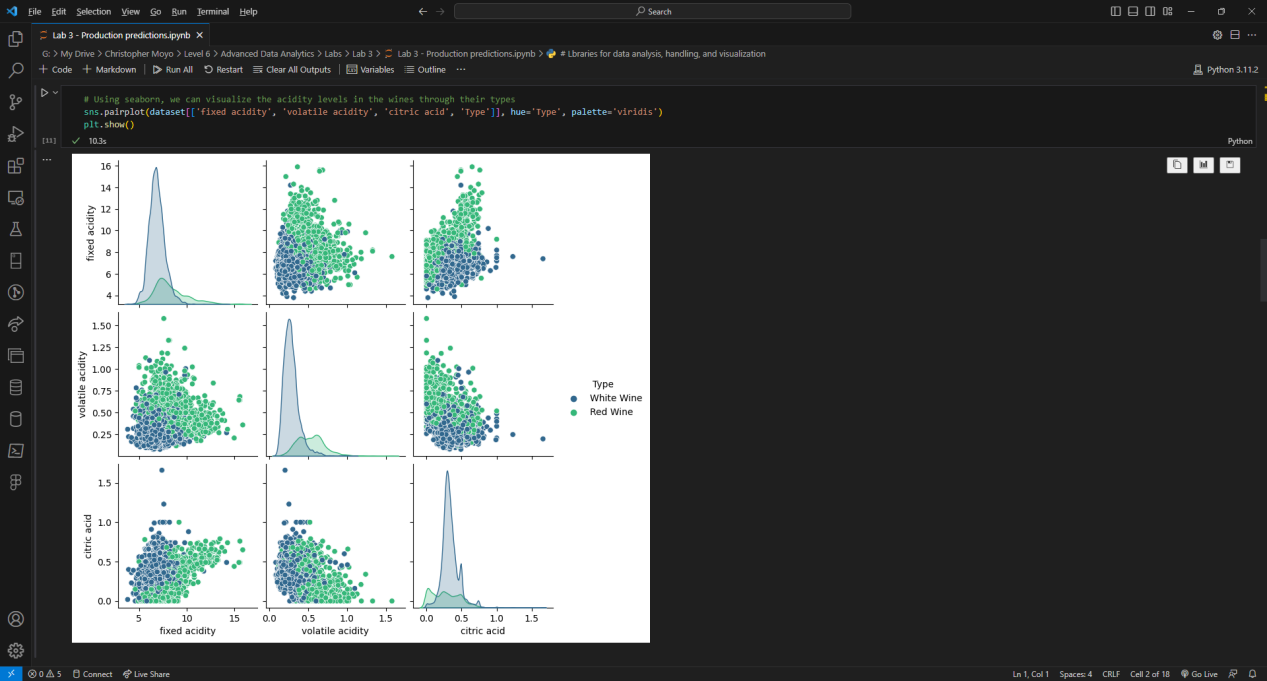
* Cells 6, 7, and 8 are created in order to show dataset information and clearing off any unnamed attributes.

1. We then take notes of the counts of each type of wine produced as shown below.

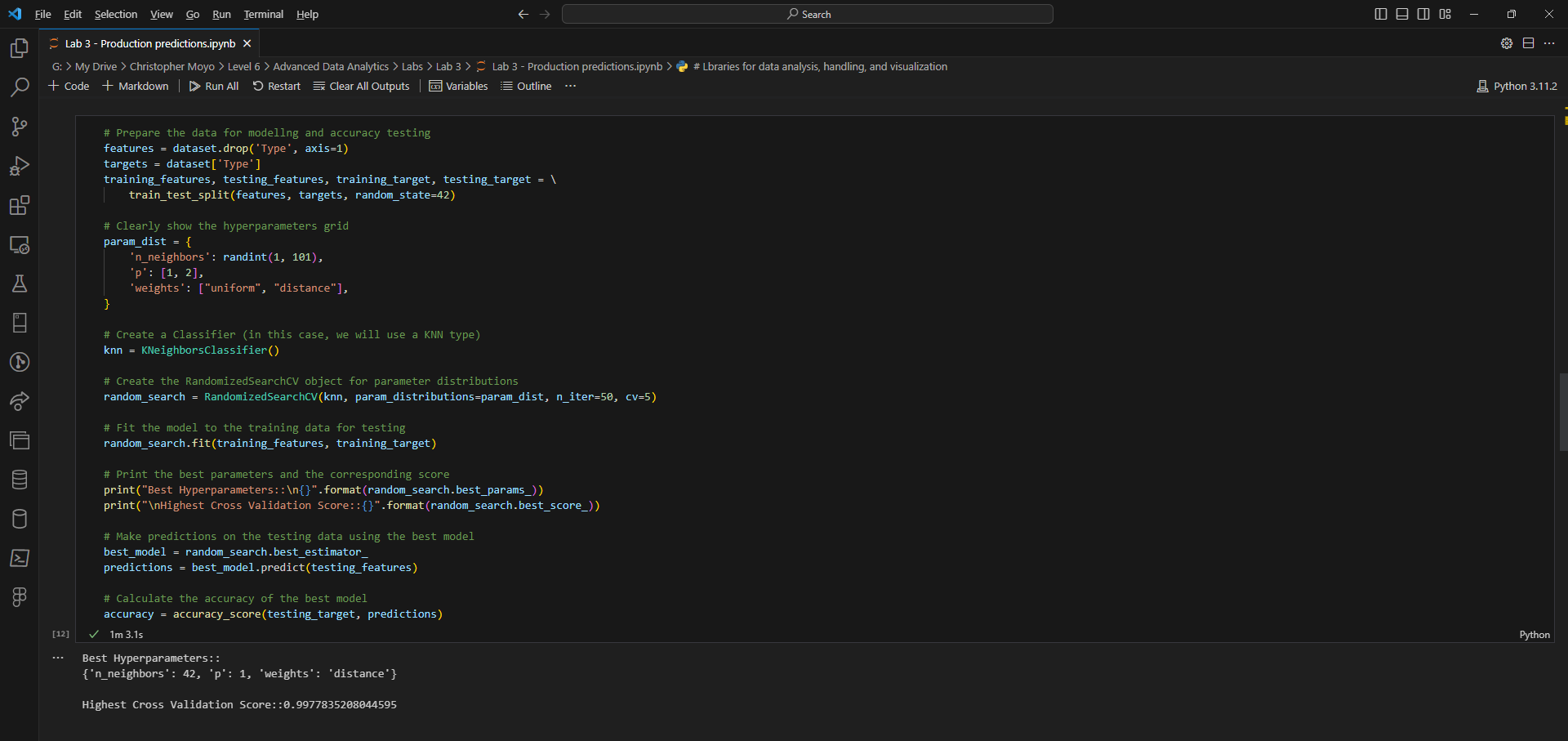


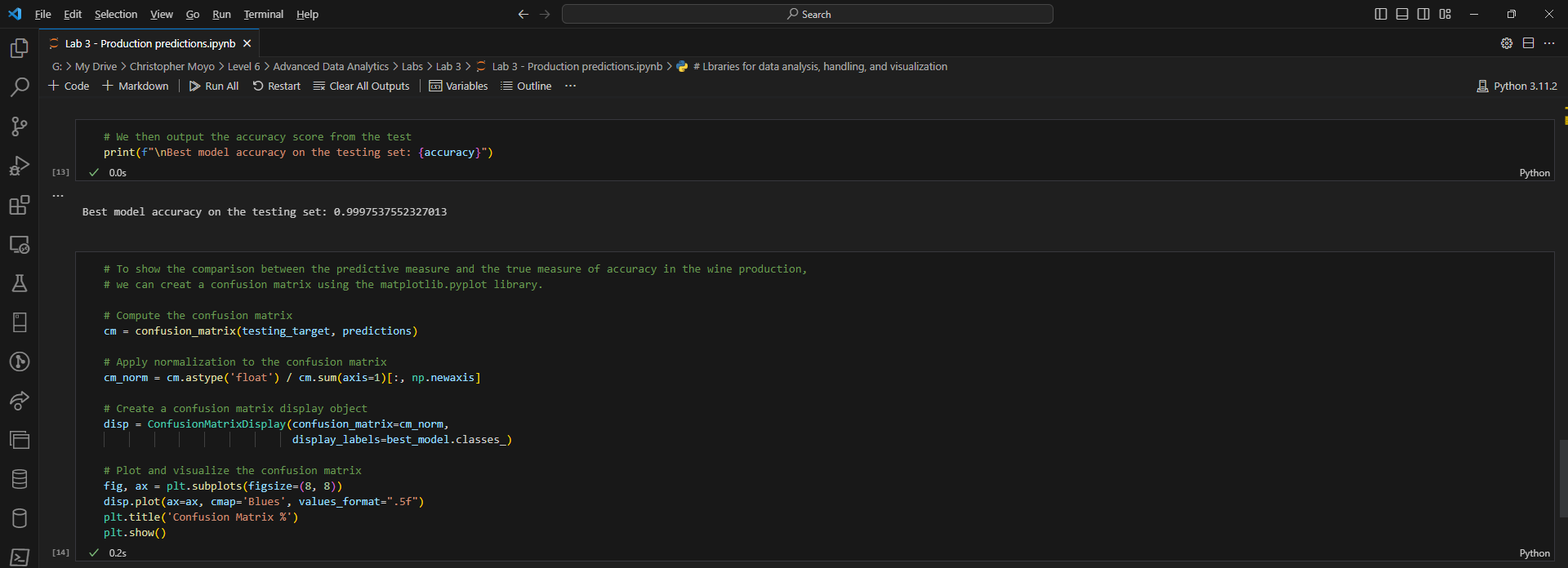
* Cells 9 and 10 are written to take note of the percentage count of each produced wine type and the quality that each quality possesses.

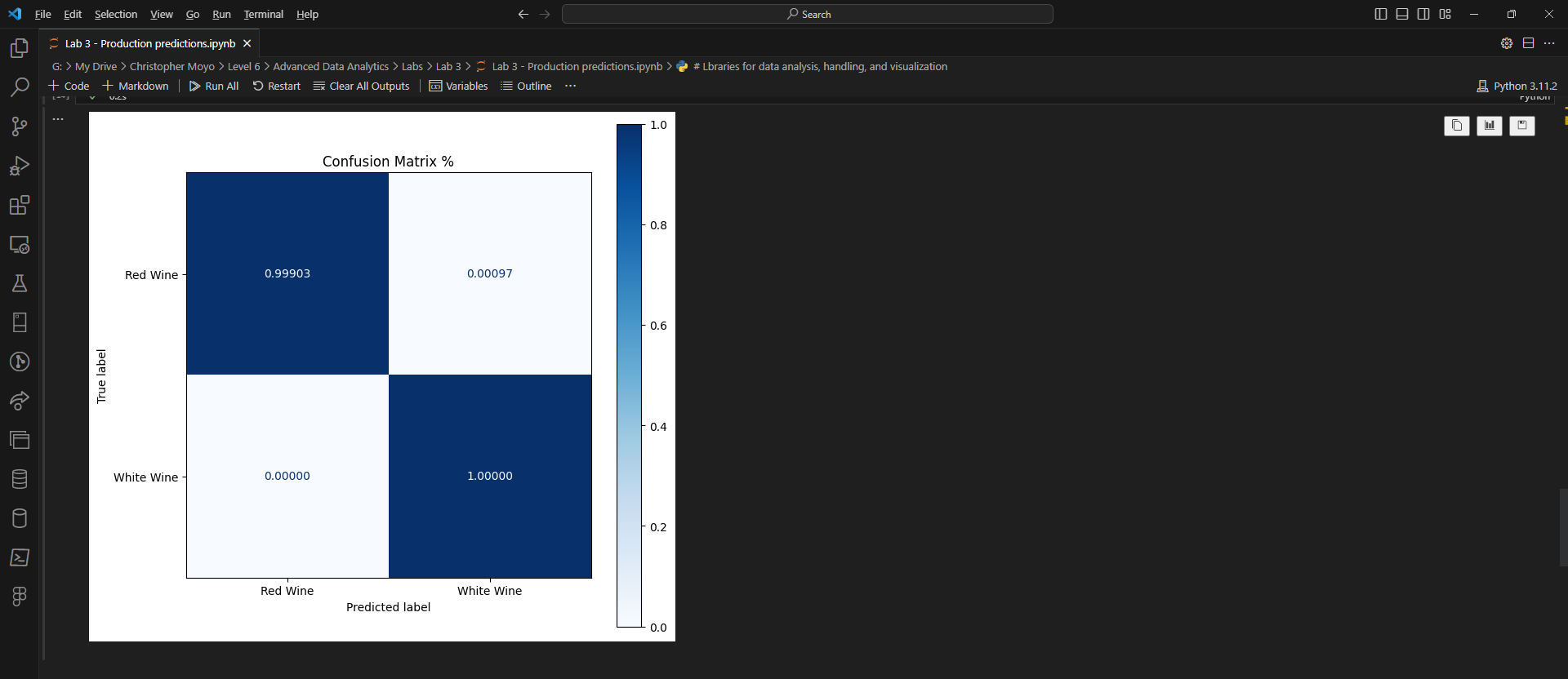
1. Using seaborn, we can visualize show the contents of the wines in terms of citric acid, volatile acidity, and fixed acidity, as shown below:

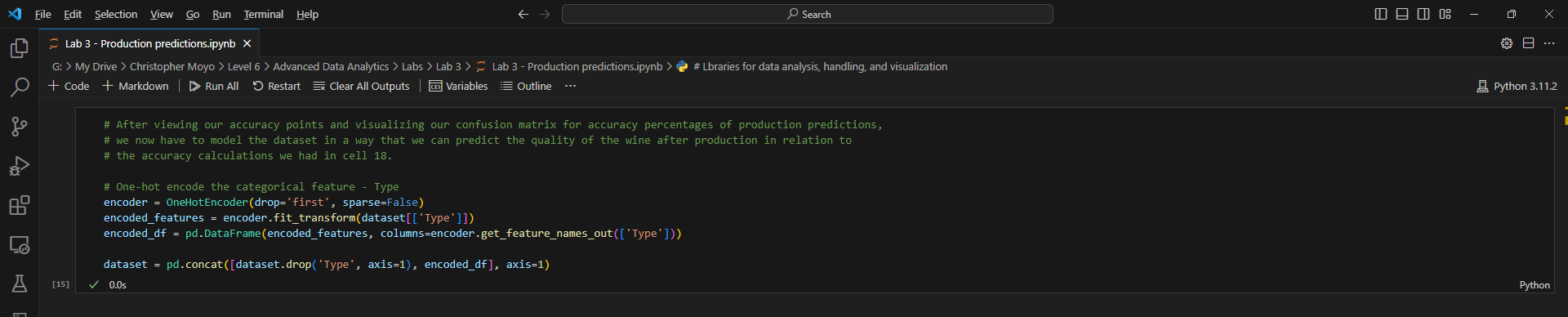


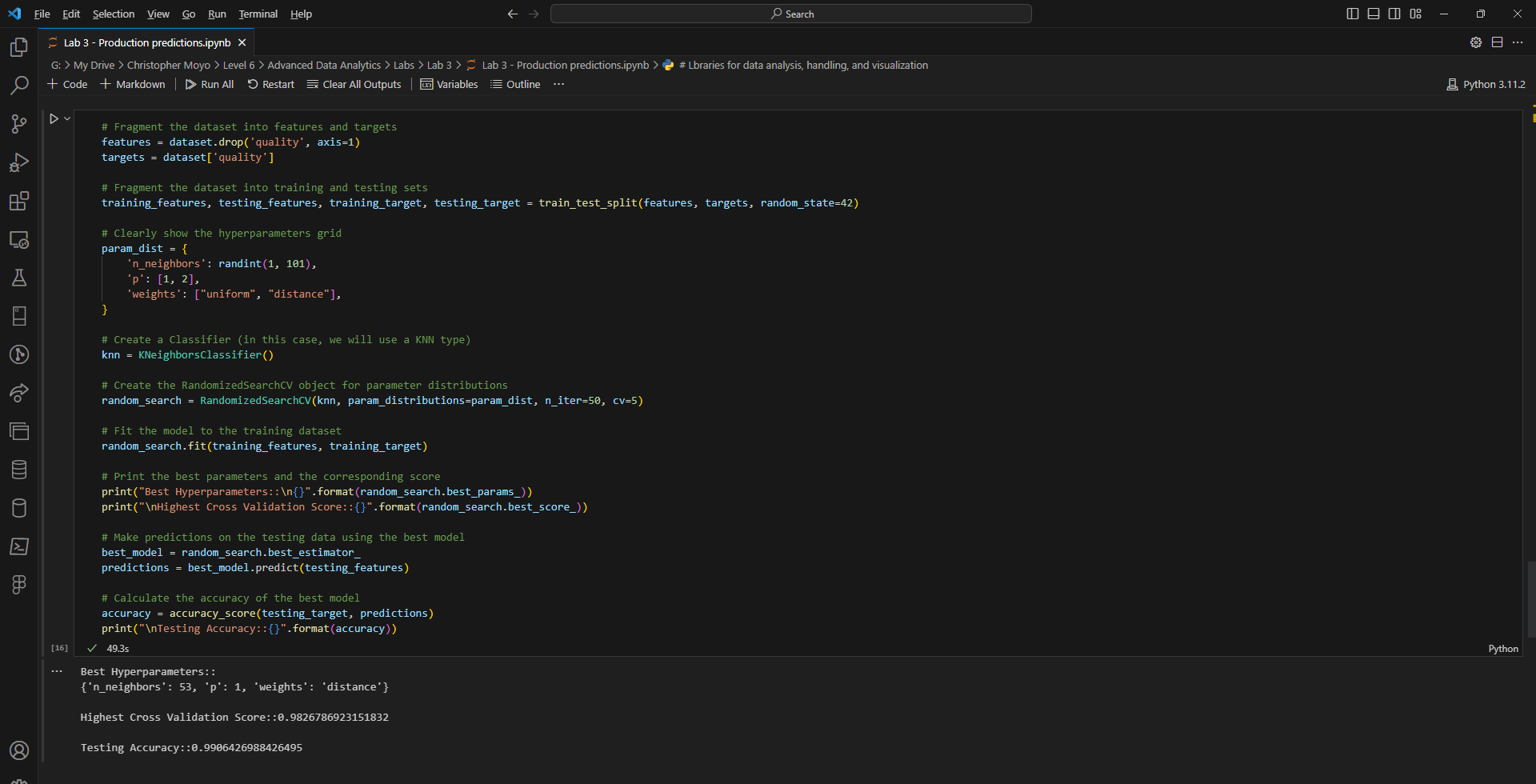
1. We then get to prepare the data using the Scikit-Learn (sklearn) library for further analysis and visualization of the wine in terms of predictive and true values of production in relation to the types of wine produced.

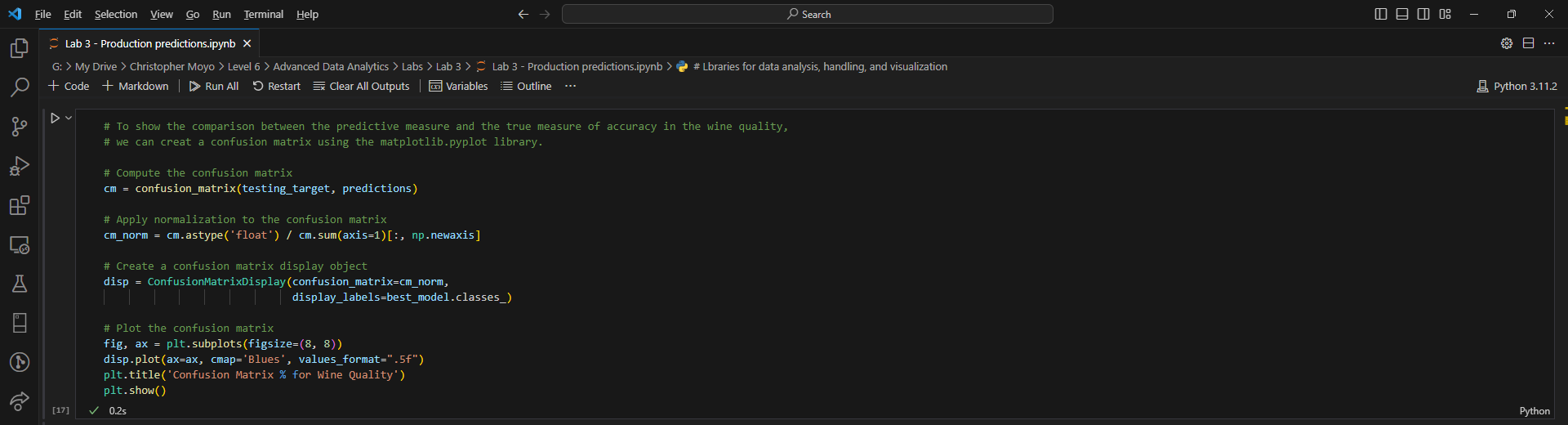


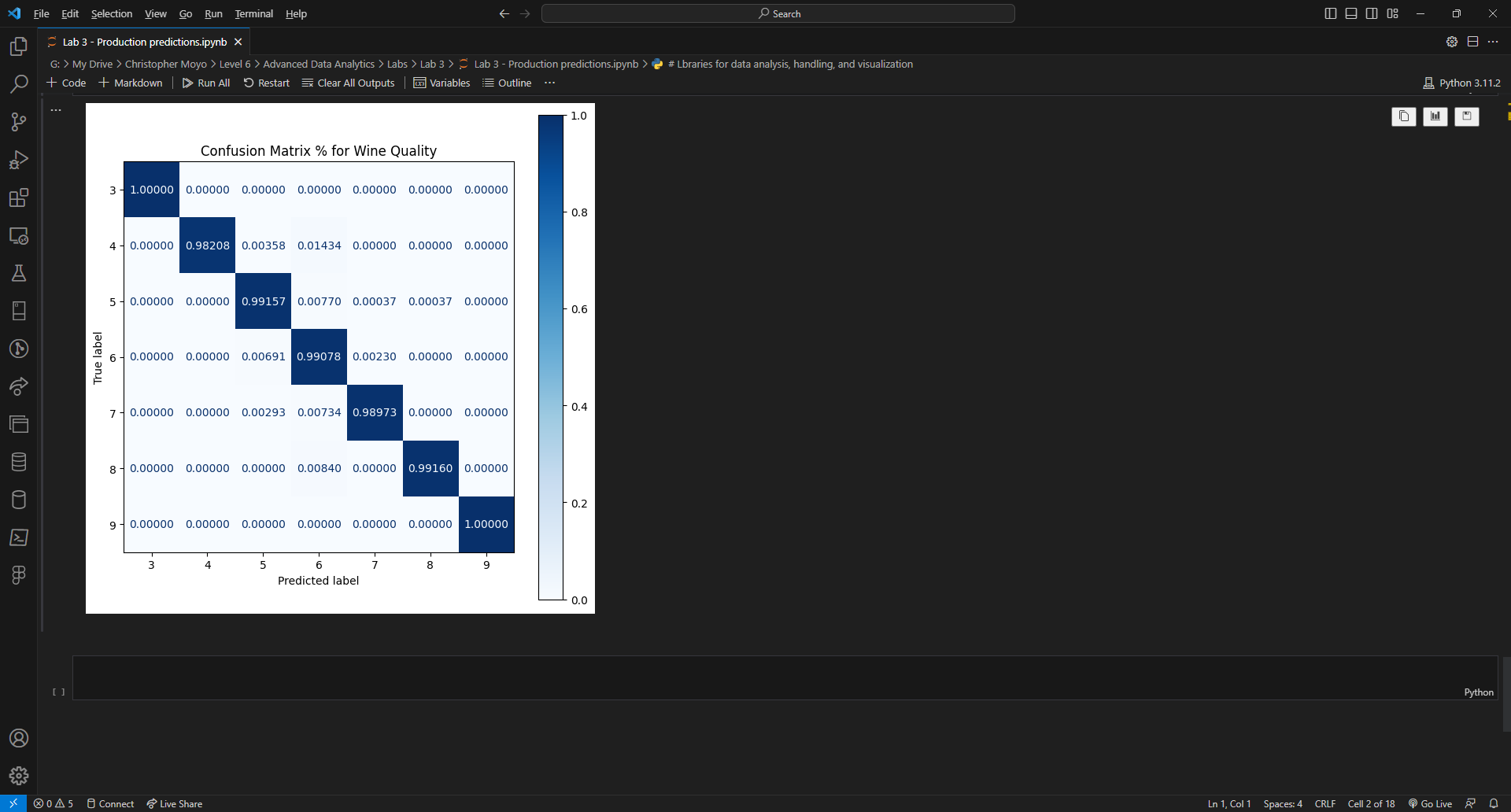












The shown graphics show the differences in predicted quality vs the true quality produced and this can be followed in the comment scripts in green writing on every screenshot above.

**References.**

1. Cyber Cop. (2023). *Wine Quality Data (Combined)*. Available at: <https://www.kaggle.com/datasets/subhajournal/wine-quality-data-combined.> [Accessed 25 August 2023].