

Họ tên: Nguyễn Hữu Hiệu
MSSV: 20520506

WINE

1. Code book

Thông tin	Nội dung
Tên bộ dữ liệu	Wine Data Set
Nguồn thu thập	S. Aeberhard, D. Coomans and O. de Vel, Comparison of Classifiers in High Dimensional Settings, Tech. Rep. no. 92-02, (1992), Dept. of Computer Science and Dept. of Mathematics and Statistics, James Cook University of North Queensland.
Số thuộc tính	13
Số mẫu dữ liệu	178
Thông tin các thuộc tính	1. Alcohol 2. Malic acid 3. Ash 4. Alcalinity of ash 5. Magnesium 6. Total phenols 7. Flavanoids 8. Nonflavanoid phenols 9. Proanthocyanins 10. Color intensity 11. Hue 12. OD280/OD315 of diluted wines 13. Proline Class 1: 59 Class 2: 71 Class 3: 48
Thông tin tác giả	Forina, M

2. Raw data

Bao gồm **wine.data**, **Index** và **wine.names**.

3. Tidy data

Tidy data được lưu lại thành file **wine.csv**

4. Instruction list

```
import pandas as pd

# Insert the directory path in here
path = 'wine.data'
```

```
col_name = ["Class", "Alcohol", "Malic acid", "Ash", "Alcalinity of  
ash", "Magnesium", "Total phenols", "Flavanoids",  
            "Nonflavanoid phenols", "Proanthocyanins", "Color  
intensity", "Hue", "OD280/OD315 of diluted wines", "Proline"]  
  
# Add column names into dataset  
df = pd.read_csv(path, names=col_name)  
  
print('Task finished!')  
  
df.to_csv('wine.csv', index=False)
```