Mục lục

[I) File: winequality-red.csv 2](#_Toc364076569)

[i) Tập dữ liệu được thu thập dùng để làm gì? 2](#_Toc364076570)

[ii) Có tất cả bao nhiêu mẫu trong tập dữ liệu? 2](#_Toc364076571)

[iii) Mỗi mẫu có bao nhiêu thuộc tính? 2](#_Toc364076572)

[iv) Thống kê thuộc tính: 2](#_Toc364076573)

[II) File: winequality-white.csv 2](#_Toc364076574)

[i) Tập dữ liệu được thu thập dùng để làm gì? 2](#_Toc364076575)

[ii) Có tất cả bao nhiêu mẫu trong tập dữ liệu? 2](#_Toc364076576)

[iii) Mỗi mẫu có bao nhiêu thuộc tính? 3](#_Toc364076577)

[iv) Thống kê thuộc tính: 3](#_Toc364076578)

BÀI TẬP 1 – HIỂU DỮ LIỆU

Thông tin nhóm:

Đàm Thiệu Quang - 1241393

Nguyễn Thị Yến - 1241444

# File: winequality-red.csv

# Tập dữ liệu được thu thập dùng để làm gì?

Tập dữ liêu được thu thập dùng để tìm các thành phần quyết định chất lượng của rượu vang đỏ.

# Có tất cả bao nhiêu mẫu trong tập dữ liệu?

Số mẫu: 1599

# Mỗi mẫu có bao nhiêu thuộc tính?

Số thuộc tính: 12

# Thống kê thuộc tính:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tên thuộc tính | Kiểu dữ liệu | Trung bình/số giá trị phân biệt | Phương sai/Số giá trị duy nhất | Số mẫu bị thiếu |
| fixed acidity | Numeric | 8.32 | 3.031 | 0(0%) |
| volatile acidity | Numeric | 0.528 | 0.032 | 0(0%) |
| citric acid | Numeric | 0.271 | 0.038 | 0(0%) |
| residual sugar | Numeric | 2.539 | 1.988 | 0(0%) |
| chlorides | Numeric | 0.087 | 0.002 | 0(0%) |
| free sulfur dioxide | Numeric | 15.875 | 109.412 | 0(0%) |
| total sulfur dioxide | Numeric | 46.468 | 1082.081 | 0(0%) |
| density | Numeric | 0.997 | 0 | 0(0%) |
| pH | Numeric | 3.311 | 0.024 | 0(0%) |
| sulphates | Numeric | 0.658 | 0.029 | 0(0%) |
| alcohol | Numeric | 10.423 | 1.136 | 0(0%) |
| quality | Numeric | 5.636 | 0.653 | 0(0%) |

# File: winequality-white.csv

# Tập dữ liệu được thu thập dùng để làm gì?

Tập dữ liêu được thu thập dùng để tìm các thành phần quyết định chất lượng của rượu vang trắng.

# Có tất cả bao nhiêu mẫu trong tập dữ liệu?

Số mẫu: 4898

# Mỗi mẫu có bao nhiêu thuộc tính?

Số thuộc tính: 12

# Thống kê thuộc tính:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tên thuộc tính | Kiểu dữ liệu | Trung bình/số giá trị phân biệt | Phương sai/Số giá trị duy nhất | Số mẫu bị thiếu |
| fixed acidity | Numeric | 6.855 | 0.712 | 0(0%) |
| volatile acidity | Numeric | 0.278 | 0.010 | 0(0%) |
| citric acid | Numeric | 0.334 | 0.015 | 0(0%) |
| residual sugar | Numeric | 6.391 | 25.725 | 0(0%) |
| chlorides | Numeric | 0.046 | 0 | 0(0%) |
| free sulfur dioxide | Numeric | 35.308 | 289.238 | 0(0%) |
| total sulfur dioxide | Numeric | 138.361 | 1806.080 | 0(0%) |
| density | Numeric | 0.994 | 0 | 0(0%) |
| pH | Numeric | 3.188 | 0.023 | 0(0%) |
| sulphates | Numeric | 0.49 | 0.013 | 0(0%) |
| alcohol | Numeric | 10.514 | 1.515 | 0(0%) |
| quality | Numeric | 5.878 | 0.785 | 0(0%) |