Group members:

Maya Hany Elshweiky 900204233

Omar Ahmed 900201077

**Description of Red Wine Quality Data**

**Data Description**

**Who:** 1599 observations of wine quality based on physicochemical tests

**When:** The data was collected in 2009

**Where:** The data set was obtained from kaggle.com and the Citation for the data set is found below:

**Source** : <https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009>

**Citation:** P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009.

**What:** There are 13 variables in the data, for each we have the wine quality (scored between 0 and 10), the score of wine quality (0 or 1) and eleven chemical attributes (quantitative), which are as follows: Fixed acidity, Volatile acidity, Citric acid, Residual sugar, Chlorides, Free sulfur dioxide, Total sulfur dioxide, Density, PH, Sulphates, and Alcohol and their description is shown below in the table.

| **Variable** | **Type** | **Unit of measurement** | **Description** |
| --- | --- | --- | --- |
| fixed acidity | Quantitative (Numeric) | tartaric acid - g / dm^3 | Fixed acids, numeric from 3.8 to 15.9 |
| volatile acidity | Quantitative (Numeric) | acetic acid - g / dm^3 | Volatile acids, numeric from 0.1 to 1.6 |
| citric acid | Quantitative (Numeric) | g / dm^3 | Citric acids, numeric from 0.0 to 1.7 |
| residual sugar | Quantitative (Numeric) | g / dm^3 | residual sugar, numeric from 0.6 to 65.8 |
| chlorides | Quantitative (Numeric) | sodium chloride -  g / dm^3 | Chloride, numeric from 0.01 to 0.61 |
| free sulfur dioxide | Quantitative (Numeric) | mg / dm^3 | Free sulfur dioxide, numeric: from 1 to 289 |
| total sulfur dioxide | Quantitative (Numeric) | mg / dm^3 | Total sulfur dioxide, numeric: from 6 to 440 |
| density | Quantitative (Numeric) | g / dm^3 | Density, numeric: from 0.987 to 1.039 |
| pH | Quantitative (Numeric) | 0-14 scale  (0 being most acidic, 14 most alkaline) | pH, numeric: from 2.7 to 4.0 |
| sulphates | Quantitative (Numeric) | (potassium sulphate - g / dm3 | Sulfates, numeric: from 0.2 to 2.0 |
| alcohol | Quantitative (Numeric) | % by volume | the percent alcohol content of the wine, numeric: from 8.0 to 14.9 |
| quality | Categorical | 0-10 scale | Wine quality score between 0 (very bad) and 10 (very excellent) |
| score[[1]](#footnote-0) | Categorical (binary) | - | - Score 1 represents good quality wine in which the quality rating is greater than 6.5  -Score 0 represents bad quality wine in which the quality rating is less than 6.5 |

# 

1. Note: score variable was added using excel to split quality rating into two categories good and bad [↑](#footnote-ref-0)