Datasets for lab 3:

<https://www.kaggle.com/szamil/who-suicide-statistics>

classify what age group in most risk

<https://www.kaggle.com/goldenoakresearch/us-household-income-stats-geo-locations>

classify range of income

<https://www.kaggle.com/christophercorrea/dc-residential-properties>

classify range of price

<https://www.kaggle.com/jsphyg/weather-dataset-rattle-package>

rain fall in Australia. Group the num of mm per day, or temp at 9 /3 of humidity

<https://www.kaggle.com/giovamata/airlinedelaycauses>

group delay time. depDeley

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

The dataset token from: <https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009>

Originally, it was posted on the UCI machine learning repository: [*https://archive.ics.uci.edu/ml/datasets/wine+quality*](https://archive.ics.uci.edu/ml/datasets/wine+quality)

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.**

The dataset related to red variants of the Portuguese "Vinho Verde" wine. It contains 11 features: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol percentage. The target of the set is the quality of the wine, which is on scale of 0 to 10. The measurement of the target was measured by sensors, which require several tests in a special lab and most importantly can be made just after making the wine.

The model I am claiming to build should use the features given to predict the quality of a wine before making it. the classification task is the scale of quality from 0 to 10. Based on the model, the wine producer should provide the required features for making the wine, and the model will return the predicted quality of the wine that will be made. Using the model, the producer will not need to wait until the end of the process of making the wine in order to measure the quality of the wine, but they will have a prediction even before starting the process.

The parties which will be interested in such a model will be the wine producers as well as the wine stores that will be able to have tool to measure the quality of the wine without the lab testing process.

For the model to be useful, it should have accuracy of at least 90%. In this case, the producers should use the model before the process of making the wine in order to set the features for best results, and a t the end still make a quality checking for more accurate testing.

MSE:

<https://towardsdatascience.com/why-not-mse-as-a-loss-function-for-logistic-regression-589816b5e03c>