

Statistical Analysis. Pandas.

Dataset: Red Wine Quality

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

One of the main goals of classification machine learning models is to try to find patterns that will allow the algorithm to separate the classes of samples. Usually, they do that using different approaches.

People can do that too ! and as a data scientist, it is important to find manually the patterns that will allow a Machine Learning algorithm to do that. Usually, we use that Statistical tools and methods.

In this task, I invite you to find patterns and features that best separate the Higher quality wine than the lower quality.

The tasks:

- Create a jupyter notebook with a clean code.
- Study some columns together and try to make conclusions in relation with the quality column. I highly recommend you to explore the following statistical parameters:
 - Mean.
 - Median.
 - Mode.
 - Standard Deviation.
 - Variation.
- Find what columns have the highest absolute correlation with the Quality column, and try to explain why do you think so happens. (here you can explore Red Wine Quality literature.)
If the correlation indexes are low you can take a range, for example > 0.1 or > 0.05 .
- Please comment on your code.