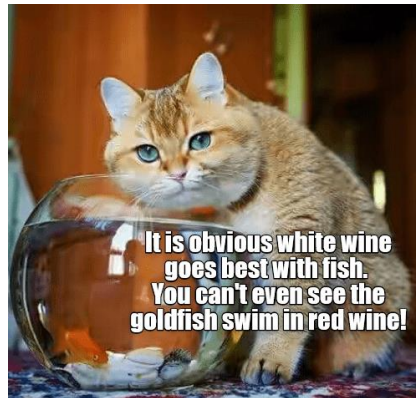


## Assignment 3 – Maths for Machine Learning Part 2



**Objective:** Applying different statistical and probability related operations to Machine Learning dataset

**Step 1: Download** the **White Wine Dataset** from <https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-white.csv>

The dataset description can be found in:

<https://archive.ics.uci.edu/ml/datasets/Wine+Quality>

**Step 2: Read** the dataset and store it as a 2D NumPy matrix (4898\*12) called XY.

**Step 3: Compute** the following statistical values (using built-in functions) for all the columns (i.e. 12) of the dataset XY.

1. Mean
2. Mode of the last column (i.e. quality of wine)
3. Standard deviation
4. Variance

**Step 4:** Perform the following operations:

1. **Check** whether any missing values are there in the dataset.
2. **Find** maximum and minimum values of each column
3. **Count** number of samples for each value of wine quality (discrete values from 1 to 10) and plot it as pie chart.
4. **Plot** the columns using bar graphs or histogram graph to show the distribution of the data.

**Step 5:** Discover the real applications of the operations used in step 3 and 4.

**Suggested Package:** numpy. **Marking:** Marking is based on both **performance during the lab hours** as well as **complete submission**.