**Feature Extraction - Wine Workshop**

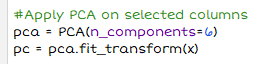
**Objective:** To practice PCA using scikit-learn (sklearn) library (<https://scikit-learn.org/stable>)

**Data**

The dataset provided is called ‘wine.csv’ and the description of the data can be found here - <https://archive.ics.uci.edu/ml/datasets/wine>

**Your Tasks**

1. Use Pandas to read in ‘wine.csv’.
2. Select all the columns, **except** the first column (Cultivar), to perform PCA on. ‘Cultivar’, in this dataset, is a label (or a class), hence we do not use it when computing our principal components.
3. Apply PCA on the selected columns with sklearn’s *PCA* function. Feel free to experiment with **n\_components**, which tells the *PCA* function the number of principal components to generate from our data.



The variable **pc** is a 2D array that contains values for our principal components. If our **n\_components** is 6, then the function would return values for 6 principal components. Each row in **pc** corresponds to the same row in the original dataset. For example, the first row (of 6 values) in **pc** corresponds to the first row (of 14 values) of our dataset.

The documentation for sklearn’s PCA function can be found here - <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>

1. Print out the **proportion** of **Explained Variance** that each principal component contributes. Then sum up the ratios and print out Explained Variance for your **n\_components** of choice. You can use the member variable **explained\_variance\_ratio\_** from the **pca** object to print out the ratios.

An example of the outcome:



1. Using Matplotlib, plot a 3D graph with the values from **Principal Components 1, 2 and 3**; denoting each sample (i.e. each row) with a color that represents one of the **cultivars** (i.e. column ‘Cultivar’ of our dataset). Label your graph’s axes and display the legend. Finally, title your graph as “PCA on wine dataset”.

Here is an example:

