## Introduction

Machine learning refers to a vast set of tools for understanding data. These tools can be classified as supervised or unsupervised. Broadly speaking, supervised machine learning involves building machine learning models for predicting, or estimating, an output based on one or more inputs. Problems of this nature occur in fields as diverse as business, medicine, astrophysics, and public policy. With unsupervised machine learning, there are inputs but no supervising output; nevertheless, we can learn relationships and structure from such data. To provide an illustration of some applications of machine learning, you are provided with real-world data set that is considered in this assignment.

**Outline**

A beverage company produces some of their beverages locally while some are imported. The experts assign each beverage a score which is an indication of the beverage quality. A score ranges from 1 to 10 where 1 is the lowest quality and 10 is the highest quality. According to the experts, a beverage with a score in the range of 1 to 6 (inclusive) is regarded as a low-quality beverage while those with a score in the range of 7 to 10 (inclusive) is regarded as a high-quality beverage.

**Aim**

* Given two datasets - local-beverages.csv and imported-beverges.csv, your first aim is to train models with these datasets such that the models can be used to classify beverages as imported or local.
* Your second aim is to evaluate and report the performance of the implemented models.

**Skills Tested**

* Basic Python Programming
* EDA
* Use of sklearn Library
* Building a classification Model and applying cross-validation
* Understanding and using different evaluation techniques
* Reporting

**Instructions and Directions**

* In the Notebook, your code here or your answer here need to be deleted and replaced with your code or answer
* You may use more than one cell to answer a question
* Ensure that you include appropriate comments
* If your code requires special instruction or dependencies to run, please provide these in your submission

## Mark Allocation (95 marks)– see the notebook

## Additional Information

* All work must be done on your own.
* Belgium Campus have software that can **scan for plagiarism** and a student caught doing this will get 0 for this assignment.
* Late assignments will not be accepted; missing the deadline is an automatic 0.
* Please submit both the pdf and notebook versions