**Practical Assessment: Programming**

**Data Quality Analysis as a Precursor to Machine Learning Activities**

A student-information-system (SIS) intends to include additional ML-based insights into its management reporting suite. One of the csv data files used to prime the SIS Machine Learning application has been found to have significant data quality issues.

You have been asked to provide advice and guidance on the steps that need to be taken to clean the data file as part of quality assurance for the ML system.

**To complete this assignment:**

Carry out the following tasks

1. Perform detailed analysis of data quality for the **dataset** provided, making use of some of the criteria discussed in the week’s lectures.
2. Map out, in outline form, what data cleansing measures would need to be put in place to prepare the dataset for ML work.
3. Indicate the types of Pandas-Scikit-Learn python commands that could be used whilst carrying out the measures you suggested in task 2.
4. Attempt a clean-up of the dataset using the procedures set out in (2) and (3) and report on the degree of success with carrying this out.

Submit the final document (in MS-Word or pdf format) containing your responses for sub-tasks 1, 2, 3 and 4 in form of a brief **report**(500 words). The **Python code** must be included in your submission.