**Key elements of the report**

**Title**

The title should provide an overview of the focus of your problem and the expected solution.

**Introduction**

This section contains a brief background to the topic and leads to the formulation of the specific question, based on your selected topic. The research question must be focused and clear.

**Datasets**

You are welcome to choose any datasets that interest you, and that has enough data to enable meaningful analysis. In making your choice, you should be sure to consider what problem or problems you would be able to solve by employing data mining on the dataset. In other words, you should ask yourself: How could I use data mining to answer one or more questions about the datasets?

**Explanation and preparation of datasets**

Briefly describe the datasets you have used, independent and dependent variables. Explain any preparation tasks (e.g., normalisation, dealing with missing values, handling class imbalance etc.) carried on the datasets.

**Implementation in Python / Azure Machine Learning Studio**

Implement your proposed approach using libraries available in Python. This section will include:

* A brief description of the algorithms used.
* The application of data-mining techniques to selected datasets that you choose using Python (or Azure Machine Learning Designer for Task 1b).
* Explanation of the experimental procedure, including the setting and optimisation of model hyperparameters during training, and your approach to validation (for supervised learning tasks).
* Visualisation of the results.

**Results analysis and discussion**

* Explain and justify the performance metric you choose to use to evaluate the model(s).
* A clear and compelling presentation of the results that you obtain, both from the data mining and any other analysis that you may perform.
* For tasks that require you to use more than one algorithm, you should compare and discuss the results obtained from each.
* You should also consider and discuss any ethical, legal or professional considerations in using machine learning and data mining on the datasets you have selected.

**Conclusions**

The key points from the assignment must be synthesised within the conclusion. This must relate back to the introduction and the research question and provide an overall evaluation of the validity of the solution you have proposed.

**References**

You will list all publications referenced in the report. You should show evidence of sufficient readings related to your work. References must follow the Harvard formatting system as in this guide:

<http://www.salford.ac.uk/library/help/user-guides/general/Bibliographic-Citations-APA-QuickRef-Apr2015.pdf>

**Appendices**

Appendices may be used to provide relevant supporting evidence for reference but should only be used if necessary. Students may wish to include in appendices, evidence which confirms the originality of their work or illustrates points of principle set out in the main text.