**Data Study Group Report Reviewer Guidelines**

DSG reports are produced by participants to document their work and main findings during the weeklong event. The core content of DSG reports is created during the 3.5 days that participants spend working on the challenge. Given the short duration, the goal of a report is to produce clearly presented preliminary insights into an important problem. Reports are key outcomes of the DSG and provide a tangible output for our organisations and participants. They are published by the Turing and form the basis for any follow up research.

Before reports can be published on the Turing website, they must be reviewed to ensure that they meet acceptable standards of scientific exposition and quality. The two key considerations are: **clear & engaging** exposition (e.g., clear description of data sets, questions asked, preliminary insights, and promising future work) and **honest caveats** for the research undertaken (e.g., limited sample size, need for further assessment of generalisability and robustness of results, etc). The right balance between these two is what makes a good report. We give further concrete guidelines on each later in this document.

Please allow between 1 – 2 hours to review a DSG report. Please feel free to amend the document as you go for small & easy fixes, such as issues with punctuation and grammar. Please do not re-write any part of the report. If more serious issues arise (the above key considerations can serve as high-level guiding principles for what constitutes a “serious issue”), highlight the issue and leave a comment. In such cases, reports will be returned to their writers so that the necessary adjustments can be made.

Please see the ‘reports’ section of the Data Study Group webpage for some example reports:

<https://www.turing.ac.uk/collaborate-turing/data-study-groups>

**Content Checklist:**

* Correctness/ Accuracy of statements (to the best of your knowledge)
* Clarity in writing and argumentation
* Clear logical explanations
* Correct spellings and grammar, with The Alan Turing Institute capitalized correctly.

**Layout Checklist:**

* An easy to read summary paragraph followed by an executive summary at the start. It should be technically correct, but readable by a non-data-scientist.
* A section which explains in technical detail how the data scientific approaches were obtained from the domain questions, and what these are.
* A section which explains the data tables that were available: for each table, sample size, rows (samples), columns (variables).
* A longer section on limitations.
* A section on future work – what would the group suggest to do if follow-up happens?
* Tables, Figures should be in manuscript (not in appendix), have full descriptive captions, legends if necessary.

**Things to watch out for:**

* **Incomplete argumentation**
* Obviously, one should avoid using “obviously”, “clearly”, or similar formulations. Because if it is obvious, it should be easy to state the full argument instead of making the reader guess.
* Make sure that the use of all methods/algorithms is justified. Formulations that imply that intimidate the reader by implying they should know why something is the only choice are a “don’t”, and should be replaced by a full argument why an approach was chosen.
* **Unsupported claims**
* Claims that are not corroborated by facts and empirical findings, e.g., that an approach shows a lot of promise, results are excellent, etc, without proper experiments.
* Claims of causality when the data does not support it – e.g., when there is no intervention or controlled trial set-up.
* **Ignoring prior work**
* Off-shelf ideas should be credited appropriately. Novel ideas should credit precursor ideas appropriately. Usually, an overview reference is sufficient, but not crediting is a “don’t”.
* **Side-stepping peer review for original methodology research**
* DSG reports are not the correct venue to place original methodology research.   
  All methodology should be in the context of the challenge, therefore proposing a novel methodology for a general case should happen in a separate publication.  
  Research inspired by challenges is a desired outcome, but of course best scientific practice should be adhered to – including peer review and crediting of the ideas landscape.
* **Marketing language and conflicts of interest**
* Reports should not market a specific product, method, algorithm. Avoid marketing language.   
  Explain why a solution was chosen, e.g., starting at requirements.
* Avoid praise of challenge owner, or mention of products or facts unrelated to the data study challenge. Generally, as in all scientific work, neutral scientific tone should be adhered to.