# Checklist for report writing, for finalizers and challenge PIs

## “Do”-s

**General**

* Where colours are used, it should be made sure that greyscale prints are still readable.
* Spell Check SPELL CHECK Spell Check!
* If referenced, ‘The Alan Turing Institute’ should start with all capital letters.

**Tables and Figures**

* Tables, Figures should be in manuscript (not in appendix), have full descriptive captions, legends if necessary.
* All tables, figures should be referenced in-text.
* Table captions should state: what the table is about; what are rows; what are columns (for each column). If in reference to data: briefly reference what data extract this is on.
* Figure captions should state: what the figure is about; all figure elements should be explained in the caption; if coordinate system: what the x-axis and y-axis is, including units; if multiple elements, e.g., curves, these should be explained. If in reference to data: briefly reference what data extract this is on.
* For tables/figures in panels, e.g., grouped visualization, scatter plot matrix, etc. Caption should state meaning of panel ordering, explain rows, columns, etc.

**Must-have sections, section structure**

* Introductory summary paragraph/ a couple of sentences, explaining any outcomes and giving a very high-level overview so that a non-data-scientist could quickly obtain an idea of what the report documents, this can point to the executive summary for more information.
* There should be an executive summary, right at the start. It should be technically correct, but readable by a non-data-scientist.
* There should be a section which explains in technical detail how the data scientific approaches were obtained from the domain questions, and what these are.
* There should be one section which explains the data tables that were available: for each table, sample size, rows (samples), columns (variables). If there are multiple tables, it should be explained how the tables are linked. This section should also explain how data was cleaned, or extracts were prepared for subsequent experiments (with consistently used reference names if there are many).
* Optimally, there is a section on exploratory data analysis (EDA). If this was not recorded, this should be clearly stated, and full EDA suggested in future work.
* There should be a longer section on limitations. It is a good idea to split this into “data limitations”, i.e., limitations of the data with regards to the questions – with suggestions how to remedy the limitations, e.g., in acquiring more data. Limitations of the analyses as they were performed. Limitations that arise through omission, e.g., forgetting to record something – which is entirely defensible since DSG week is hectic, but as a note that it should be properly re-done when there is more time and leisure to do so – e.g., in a sub-section “results pending corroboration”, or “results not recorded”.
* There should be a section on future work – what would the group suggest to do if follow-up happens? Usually, the analyses will be not entirely complete, so think how these would be completed. It is also important to reflect on the challenge questions that were not at all addressed during the week. Similarly, think how the use case problem could be addressed with different data, if you would have it.
* Further sections (usually in the middle) should describe the different approaches. It is a good idea to not split or order these by individuals, but by scientific approach.

**Executive summary**

* Challenge questions should be concordant with those set out in the 2-page description. This needs not be identical and can be shortened, but it should not substantially differ.
* “approach”, “findings” should reference the full set of challenge questions – avoid tacitly ignoring a key question.
* If the group ended up not working on certain questions, for whichever reason, e.g., it was found that the data could not answer that, or if it was found a re-formulated question would be better: mention that explicitly, and explain – this is valuable information.

**Section 2 (or 3) – “explaining the approach”**

* There should be a section which explains how the domain questions map on data scientific approaches – this is very important work, sometimes one of the key informative outputs.
* Optimally, this should recount the full process of brainstorming, planning, and selecting of exploration avenues – all this is valuable insights generated in DSG week.
* The section should highlight problems separate from solutions – e.g., explain the supervised learning task separate from specific algorithms that solve it.
* Citations for approaches should be given. A textbook reference for a common off-shelf approach (e.g., “supervised learning”), or a peer-reviewed academic publication for a more specialized approach.
* If an approach is novel and/or experimental, inspired by the challenge, that is fine, and a desired outcome if such an approach is found necessary, as it highlights a need for research. However, if an experimental approach is used, it should be expressly highlighted. “Limitations” and “future work” sections should point out that future work and peer-reviewed publication is necessary for best scientific practice.

**Reporting of experiments/approaches**

* Some approaches – e.g., visualization, unsupervised learning, GUI design – will not fit the template “experiment” section’s format well. Check whether the structure is appropriate.
* All experiment/approach sections should state clearly which data these were run on. Reference data views/extracts from the “data” section.
* All experiment/approach sections should clearly and centrally state the purpose, the problem to be addressed, scientific hypotheses to be checked (if experimental), or the type of hypotheses to be generated (if exploratory).
* “experiments” should be sections describing the common workflow of (1) method building, (2) evaluation with respect to a task (e.g., supervised).
* The following things need to be described in a benchmark/performance estimation experiment: (i) what the data is on which it is conducted. (ii) what precisely the contender methods/algorithms are. This includes tuning. (iii) of the methods/algorithms, which are uninformed baselines, which are state-of-art methods (if applicable). (iv) how exactly test/training sets/splits are selected, if applicable. (v) What the performance criterion is with respect to which the methods are evaluated. (vi) how confidence intervals were computed (if applicable). (vii) how precisely methods were compared.
* If any of the above is unknown, either state clearly (e.g., “limitations of analyses” or “results pending corroboration”),

**Acknowledgments and crediting**

* All contributors should be acknowledged with a short description of their contributions.   
  Usually that will include, at least, participation in discussion and brainstorming.
* Contributors may include participants of the challenge, participants of other challenges who contributed ideas in discussion during the week, challenge owner delegates, the challenge PI, organizers, and, in rare cases, DSG external individuals with a coincidental contribution (the last need to be reviewed by DSG team due to data sensitivity and contractual issues).
* The challenge PI should also be credited with preparing and guiding the challenge, even if not present during the week (assuming the challenge PI indeed contributed in this way).
* People who have not contributed scientifically should not be acknowledged.
* DSG organizing team members should not appear with individual acknowledgment, unless they contributed scientifically. By default, organizers appear via consortium acknowledgment (“DSG”) that acknowledges their general roles.

## “Don’t”-s

**Bad science: 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.

**Bad science: unsupported claims**

* Avoid claiming anything that is not corroborated by facts and empirical findings, e.g., that an approach shows a lot of promise, results are excellent, etc, without proper experiments.
* Avoid “reading the entrails” of a piece of visualization – i.e., avoid claiming of trends, patterns, conclusions, by isolated inspection of a picture, without proper quantification (e.g., frequentist hypothesis test). This “don’t” is often found combined with the use of “obviously”, “clearly”, or similar.
* Avoid “reading the entrails” of model parameters – i.e., avoid claiming connections, relations, without appropriate quantification (e.g., frequentist hypothesis test). The most frequent instance of this “don’t” is reading something into a correlation or model coefficient of 0.001 or similar because it supports some personal theory.
* Avoid claiming causality when the data does not support it – e.g., when there is no intervention or controlled trial set-up. In causal models for observational data, make the assumptions (e.g., no unobserved confounders) very clear and discuss how plausible they are. List these in “limitations”.   
  Implying that you have found “causality” (or, “root cause drivers”, “influential variables”, or however one wants to call it) without due diligence is a “don’t” (and malpractice).
* Avoid to make strong recommendations unless they can be justified. Usually, the limited nature of a DSG study only allows for recommendations on data collection or future study.

**Bad science: 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”.
* If prior work on the data was done, or there are common approaches in the field, these should be mentioned – and, if available or feasible, compared to.

**Bad science: 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.
* A common way to avoid side-stepping the academic review process is to state novelty clearly, and reference an external research publication (e.g., on arXiv) which focuses on one particular novelty.
* As a corollary, one should avoid claiming that less well-known, or speculative methodology is the “standard” for a specific application.

**Bad science: marketing language and conflicts of interest**

* Don’t 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.