# Report Guidelines

For your report, you can use the Quarto template we provide here. Please note that the Personal Information And Non-Plagiarism Statement indicated below must be integrated into the final report during your submission.

## Why Do You Write A (Technical) Report?

The aim of writing a report is to communicate results from a study to someone who did not participate to it.

The report should be relevant, **easy** to read and **convincing**. Advices:

- Take some time to ask yourself what the aims of your report are, and to whom it is addressed.
- Ask yourself which messages you report should convey and which is the best way to convey it.

## Level Of Details

Including all the minutes of your research is NOT interesting. **Relevance** and **reproducibility** are the key words here. On the one hand, the report should include all the needed details for the reader to be able to repeat on his own the relevant part of the study. On the other hand, the report is not a course about a given subject. In particular, it is not expected that you include in the report all the theory about a method that is used. Of course, it is not easy to find the good balance, which anyway probably varies from one reader to another. Often, this contradiction can be solved by referencing to clear and accessible sources (paper, book chapter, webpage, etc.).

## The Structure

There are several ways to write a report. The following structure can be used and adapted according to the specificity of the study.

### A. Title Page (Mandatory)

One separate page. It should include

- Title: large format. It should be short AND informative. Never more than a sentence.
- Logo: UNIL/HEC at least; optional: logo of the company
- Date of edition
- Course name and the year
- Group name

### B. Personal Information And Non-Plagiarism Statement (Mandatory)

One separated page containing

- Full names of all the group members and their email address
- The following text **hand-signed by all the authors**. This project was written by us and in our own words, except for quotations from published and unpublished sources, which are clearly indicated and acknowledged as such. We are conscious that the incorporation of material from other works or a paraphrase of such material without acknowledgement will be treated as plagiarism, subject to the custom and usage of the subject, according to the University Regulations. The source of any picture, map or other illustration is also indicated, as is the source, published or unpublished, of any material not resulting from our own research.

### C. Summary Or Abstract

Usually one page, it is limited to one-and-a-half-page maximum. The summary states very succinctly and clearly the issue, including a data brief description, the methods used, the main results, and the main conclusions. This summary allows someone to know, without to read your work, what has been done, why and what are the results and implications. Advises for writing it:

- Write it at the very end the work, after the report body is written and clean.
- Take each section (Introduction, etc.) and summarize it in one, two, three, sentences, one paragraph or so. Use the needed space. State the important results.
- Once finished, read it again. Cut/shorten long and useless words (e.g. "has been" -> "was", "therefore" -> "hence", etc.).
- If it is still too long, cut some part by order of importance.
- The summary is not a teaser to read the report. It contains the results. It should summary them. It should not be only inviting the reader to read the report but propose an alternative for readers who do not have time to read it all.

- Typically, "We have made an analysis of the data in this report about the influential factors on the sales. We decided to use an ARIMA model to make the forecasts." It contains a piece of information but is quite frustrating. More useful: "Using an ARIMA model, the [what] sales were found influenced by the [what] factors. Such model achieved an RMSE of [what] on forecasting a [what] test set.".

## D. Table Of Content (Toc), Table Of Figures (Tof) And Table Of Tables (Tot)

ToC is mandatory. ToF and ToT are nice to have especially for very long reports, or when these elements are central to the report. ToC must be up to date. You gain generating it automatically (easy with tools like word, LateX, Rmarkdown, etc.). ToF and ToT should also be generated automatically. This implies that figures and tables are correctly inserted with a caption.

## E. Body Of The Report

### 1. Introduction

In brief, the introduction explains why the theme is interesting, what is done in the study and what it will bring. After reading it, the reader should have a broad view on the study, why it is relevant, and how it is organized. The purpose of the introduction is not to detail all theoretical aspects. Introduction is a section that should contain

- The context and background: course, company name, business context.
- Aim of the investigation: major terms should be defined, the question of research (more generally the issue), why it is of interest and relevant in that context.
- Description of the data and the general material provided and how it was made available (and/or collected, if it is relevant). Only in broad terms however, the data will be further described in a following section. Typically, the origin/source of the data (the company, webpage, etc.), the type of files (Excel files, etc.), and what it contains in broad terms (e.g. "a file containing weekly sales with the factors of interest including in particular the promotion characteristics").

- The method that is used, in broad terms, no details needed at this point. E.g. "Model based machine learning will help us quantifying the important factors on the sales".

- An outlook: a short paragraph indicating from now what will be treated in each following sections/chapters. E.g. "in Section 3, we describe the data. Section 4 is dedicated to the presentation of the text mining methods…"

## 2. Data Description

The data description section may be optional. It is typical of report for project oriented toward statistical and machine learning. The data description section can be included as a subsection of the introduction or be a whole section by itself, depending on the complexity and the size of the database. If relevant, it should contain

- Description of the data file format (xlsx, csv, text, video, etc.)
- The features or variables: type, units, the range (e.g. the time, numerical, in weeks from January 1, 2012 to December 31, 2015), their coding (numerical, the levels for categorical, etc.), etc.
- The instances: customers, company, products, subjects, etc.
- Missing data pattern: if there are missing data, if they are specific to some features, etc.
- Any modification to the initial data: aggregation, imputation in replacement of missing data, recoding of levels, etc.
- If only a subset was used, it should be mentioned and explained; e.g. inclusion criteria. Note that if inclusion criteria do not exist and the inclusion was an arbitrary choice, it should be stated as such. One should not try to invent unreal justifications.

In some cases, data structure can be more complex like images, videos, etc. This should be adapted accordingly. Any unknown element should be mentioned as such. Overall, this description can be quite short or very long, depending on the format of the data. For example, one xlsx file with 10 features and 50'000 instances representing customer behaviors is in general quite easy to describe. If you have 25 files with different formats from different stores representing these customers, with features and instances only partially matching, then it may be more difficult to deal with. Typically, the first and simple case would be included into the introduction as a subsection while the more complex case would deserve a whole data description section.

## 3. Methods

A very important section describing in detail all the methods that are used in the report. This section explains how the study was conducted. It aims at giving relevant information and tools for others to be able to reproduce the study. It explains the global strategy of the study, and describes each tool used to achieve this goal. For a comprehensive and structured presentation, often it is important to give the general structure of the study, and then to drill down each step. The individual tools (e.g. machine learning methods, etc.) should be described in a concise and relevant way. **This section is not a course**. For example, one can give a brief definition and the important properties (important for the study), then refer to an external source for the details. In addition, the description should be balanced. Do not spend three pages on explaining the boxplots when a predictive model of random forest is explained in one-half page. It should describe and duly cite

- Statistical and machine learning methods to analyze the data

- Software, computer programs, function packages (typically for R).
- The sequence of analysis, the reason of each method.

However,

- It does not include methods that are not used, even if you spent a lot of time on it. Except maybe if the fact that a particular method is not useful is of interest for the report.
- It is **not a course** about a method. It should contain all the relevant pieces of information, but it should also be stopped at some points. Clear references and well-chosen are of major importance here.

### 4. Results

Obviously, this section presents the results: application of the methods to the data. It presents all the results (even negative ones) and only the results (no more data or method descriptions). This section should be very well structured especially if there are lot of results. Use subsections and clear references. In addition, use the appendix. For example, if 30 time series were analyzed and the corresponding 30 figures produced. Then only the most interesting ones should be included in the results section. The other ones should be reported in the appendix (or supplementary material) and appropriately cited. The interpretation can be included at this point but should be limited to technical aspects. Keep discussions for the next section.

### 5. Recommendations And Discussion

In this section, the interpretations of the results are given:

- Summary of what was observed from the results
- Implications for the business (link with the context and the initial objectives of the study).
- Limitations (this may be the subject of a separated section). An opening is often expected here: "What next?" "What can be done for improvement", etc.?

### 6. References

All external sources should be cited. The APA system is a good and safe one. One can use the Reference tool of Word. For more details, see http://www.apastyle.org/manual/index.aspx.

## 7. Appendix

The appendix section is **not a trash** in which one puts all what was done but not integrated in the body of the report. **All the elements in the appendix must be cited in the text**. Appendix can contain graphs, texts, tables, codes, etc. For example, in the body of the report: "Fig. 1 presents the analysis of the times of the process A [...]. The 29 other processes are represented in Fig. 40 to 68 in Appendix B."

## F. Supplementary Material

In this section, you mention and briefly describe all the external documents given with the report. For example, a code file name "OurSuperRCode.R" attached to the report should be mentioned here.

## Further Remarks

### General Format And Style

- Font: Times New Roman, 11pts
- Page format: A4
- Text justification: even on both side
- Margins: 2.5cm
- Max. 40 pages for the body of the report (i.e. not including title, toc, introduction, reference, appendix, sup. mat.). It is not an expected value. It is an upper bound.

### Text editor

Using Word is probably the simplest choice, but *LaTeX* is also welcomed (see Overleaf for easy and free usage) Freely available alternatives exist like LibreOffice. It should be noted that `Quarto` and `Rmarkdown` could also be a useful tool. However, be mindful of mixing research (i.e. working R codes, running complex R code or procedure) and the writing of the report. Often, one loses more time and encounter lots of difficulties. Nevertheless, `Quarto` is very valuable for writing reproducible reports. In any cases, pay attention to use cache in order to avoid running again and again very heavy computations each time you want to compile the report.

### Maths and formulae

If you need to edit equations or formulae, you may gain using an equation editor like MathWord. Free alternatives exist like LibreOffice.

**Numbers, figures and other remarks**

Usually, in the text, numbers zero (0) to ten (10) are written in letters, then 11, 12, etc. This not for reference like "in Table 1" should never be written "in Table one". Figures, tables and sections referenced in the text are names an should have a capital letter. One should write "in Table 1", "in Figure 3 shown below", "Section 3.2 shows […]". It is not the case when you refer to an object like "the next table", "in the following section we can see […]", etc.

Do not use don't (except when quoting speeches for example). For a better scientific writing, check out this blog.