

Description of the project to be submitted

Stephan.Huber@hs-fresenius.de

HS-Fresenius: Data Science for Business

Contents

Abstract	2
1 Main goal	2
2 Details	2
2.1 Dataset	2
2.2 Question	3
2.3 The report	3
2.4 The presentation	4
2.5 rmd file	5
3 Submission	5
4 Evaluation	6
5 Helpful stuff	7
Literature	7

Word count: 1304

Abstract

In the following I describe the project that needs to be submitted in the course *Data Science for Businesses*. I give some hints for your efficient progress and success, I introduce the elements and files to be submitted, and I describe how I evaluate the submissions.

1 Main goal

Course description “*Students complete this module with a project work. The project work includes a project report (15-20 pages) and a project presentation (20-30 minutes).*”

Project work

- Find an interesting dataset,
- find a question that you aim to answer with data at hand,
- try to answer the question using R and appropriate empirical methods,
- write a report of the project, and
- present your current status of the project in class.

2 Details

2.1 Dataset

- You can search for any kind of data. From my side there are no quantitative or qualitative specifications.
- If you think you have found a data set, please contact me so we can discuss whether your data is appropriate.

2.2 Question

- When you think about a question to the data, try to come up with something that you are interested in.
- Rest assured that this does not have to be a scientific question that can be answered accurately with the data at hand. An exploratory question is perfectly fine. I will give some examples in class and in some exercises.
- Again, I highly recommend to talk with me about your question and your goals.

2.3 The report

- The report should be about 5000-6000 words (which is about 15 double-spaced pages).
- Unlike an *academic paper*, this is a report in which you should *just* document, discuss, and present your project. The report should introduce your work to me. It's similar to reports that you'll have to write in business, where your boss wants to know
 - What you did,
 - why you did it the way you did,
 - what obstacles you overcame,
 - what challenges, problems and weaknesses remain, and
 - how you would suggest proceeding with your work if you would have more time and resources available.
- Please, don't try to impress me with a fancy layout or anything alike. Focus on the content and getting your knowledge across to the reader! Anything that helps with this is welcome.
- Guide and motivate the reader and outline the target audience of your work. Usually, the introduction is a good place to introduce the scope and content. In particular, make clear what is found in each section.

- Be concise. Remove all unnecessary repetition. Read each sentence several times and ask yourself if it is concise and clear and if it fits with what was said before and after.
- The paper should be written with **R Markdown**. The .Rmd file with which this document is written, can be used as a template. This file is hosted at [my GitHub page](#): Information about writing and publishing with R Markdown can be found here:
 - [R Markdown from R Studio](#)
 - [R Markdown: The Definitive Guide](#) written by Xie, Allaire, and Grolemund (2018).
 - [R Markdown Cookbook](#) written by Xie, Dervieux, and Riederer (2020).
- Insert R code in your R markdown file by typing the chunk delimiters (see the keyboard shortcut *Ctrl+Alt+I* for Windows and Linux based OS and *Cmd+Option+I* for Macintosh OS) or [this lesson](#)).
- The outline of the paper **must** contain the following building blocks:
 - Title and all common personal details (name, email, ...).
 - Word count (see this rmd file on how to give a word count).
 - Abstract of the paper (which highlights the content of the document).
 - All the R code that is necessary to replicate your results.

2.4 The presentation

- Write the presentation using R Markdown and publish it as .html and/or a .pdf file.
- Focus on the important things.
- Try to stay on time.
- Nobody is perfect and the project is done under time pressure. So don't try to sell yourself too hard. If you see weaknesses in your work, this is a good place to discuss them.
- Describe and present your data set so that everyone has an idea of the structure and

content of the data.

- Describe your question and if possible describe your plan to address the question with the data. Rest assured that you don't have to come up with a perfect strategy. It's okay if you don't know many empirical methods.
- To facilitate the organization and scheduling of all presentations, please let me know the times you are available by **October 26**. A maximum of 3 presentations per meeting would be ideal. If you do not let me know your preferences, I will determine the time and date. Here is the link where you can give me your preferences and your availability: [Doodle Poll: Data Science Presentation](#)
- We start with the presentations on October 10.

2.5 rmd file

- The rmd file should contain the complete workflow including data import, data cleaning, and data analysis.
- Not all code or output of the code has to be shown in the pdf paper. See [Code Chunks and R Markdown](#) how to set certain options to prevent code and results from appearing in the finished file.

3 Submission

- Submission deadline for academic papers and written assessments: **13 February 2023**
(please verify!)
- Upload only **one .zip file** containing the following:
 1. the paper as (a) .pdf and a (b) .html file.
 2. the .Rmd file
 3. the data set (if not too large),
 4. the presentation as (a) .Rmd and a (b) .html file,

- 5. additional files, if needed, so that I can evaluate your work.
- Please also submit your work to my Github account at <https://github.com/hubchev/dsb-2023>

4 Evaluation

- *65 % – Quality and execution of the project* – After your presentation, we will discuss your work in a personal meeting. The goal of this conversation will be that we agree on certain standards by which I will grade you. By this I mean that we define certain goals that you should achieve with your data set and your question. The goal is to create a transparent set of expectations on my part. So that you have an indication of what you need to accomplish at a minimum in order to pass the course.
- *35 % – Quality and execution of the presentation*
- I will try to evaluate your work as objectively as possible. In particular, I will
 - check whether your submission is complete, or not,
 - check whether your empirical work can be reproduced,
 - check if all formal criteria are met,
 - check for plagiarism,
 - read your work and evaluate your writing skills (clarity, coherence, grammar, etc.),
 - review and evaluate the novelty of your project,
 - evaluate the technical level of use of the programming language R for your empirical goals,
 - evaluate your argumentation,
 - assess whether your empirical reasoning makes sense and discuss your remaining weaknesses,
 - acknowledge your learning process.

5 Helpful stuff

R Markdown

To knit to all formats that are mentioned in the header, type that into the console (of course, don't forget to refer to your working directory using `setwd()`):

```
setwd("/home/sthu/Dropbox/hsf/github/courses/rmd/")  
rmarkdown::render("22-10_dsb-project-description.Rmd", "all")
```

Git and GitHub

As you should submit your work to my Github account, you can learn how to do that by following the instructions of this repository: <https://github.com/firstcontributions/first-contributions>

Literature

Xie, Yihui, Joseph J Allaire, and Garrett Grolemund. 2018. *R Markdown: The Definitive Guide*. Chapman; Hall/CRC.

Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook*. Chapman; Hall/CRC.