Description of the project to be submitted

Stephan.Huber@hs-fresenius.de

Last compiled on 26 April, 2023

Abstract

In the following I describe the project that needs to be submitted in the course *Data Science for Business*. 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.

Contents

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 academic research paper that provides the dataset that are necessary to replicate the results shown in the paper,
- replicate the results shown in the paper using R,
- write a report of your work in progress, and
- present your current status of the project in class.

2 Details

- Choose your paper wisely. It should be a compelling read for you, and you should have a basic understanding of the research methods applied. Papers can vary greatly in complexity, length, and methodological level.
- You don't need to replicate the entire paper. Replicating a single table, specific columns, or a single visualization can be sufficient. The goal is to document your progress and to demonstrate your proficiency in using R as a tool to work with data.
- Reflect on both your successes and failures, as well as any challenges you encountered. For instance, if a line of code took a significant amount of time to write, you can explain the issue to me, but keep it concise.

- I highly recommend that you discuss the paper you want to replicate with me and ask for supervision. This will help ensure that you are on the right track and have the support you need.

- I kindly request that you come to me rather than waiting for me to come to you. This will help facilitate our communication and ensure that we are both on the same page.

2.1   The report

The report should aim to be around 5000-6000 words, or approximately 15 double-spaced pages. Please note that this report is different from an academic paper in that it should focus solely on documenting, discussing, and presenting your project. Its purpose is to introduce your work to me in a way that is similar to reports written in business settings, where you would need to convey what you did, why you did it that way, what obstacles you overcame, what challenges, problems and weaknesses remain, and how you would suggest proceeding with your work if you had more time and resources.

Please refrain from trying to impress me with a fancy layout or any extraneous details. Your primary focus should be on effectively communicating your current state of work to the reader. Feel free to include anything that can help achieve this goal.

Please put some emphasize on guiding and motivating the reader. For example, the introduction is a good place to introduce the scope and content of the report.

To ensure conciseness and clarity, please eliminate all unnecessary repetition. Take the time to read each sentence multiple times and ask yourself if it is concise, clear, and coherent with what was said before and after.

Finally, please use R Markdown to write your report. You can use the .Rmd file included with this document 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, & Grolemund (2018). - R

Markdown Cookbook written by Xie, Dervieux, & 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 below 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.

2.2   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 paper and the data set used in it.
- To facilitate the organization and scheduling of all presentations, please let me know the times you are available by **April 22**. 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

2.3 rmd file

- The rmd file should contain the complete workflow including data import, data cleaning, and data analysis.

- All code but not all the output generated by the code should 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: **24 July 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,
    - check if the replication of the paper was already done with R by somebody else,
    - read your work and evaluate your writing skills (clarity, coherence, grammar, etc.),
    - review and evaluate the difficulty level of your project,
    - evaluate the technical level of use of the programming language R for your empirical goals,
    - assess whether your empirical reasoning makes sense and discuss your remaining weaknesses,
    - acknowledge your learning process.

5 Helpful stuff

5.1 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("23-04_ds-project-desc.Rmd", "all")
```

5.2 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

5.3 Word count

You can include a word count in various ways. Here are two alternatives:

This code installs and loads the required packages and save the words counted:

```r
#install.packages("devtools")
#library("devtools")
#devtools::install_github("benmarwick/wordcountaddin", type = "source", dependencies =
library("wordcountaddin")
wordcount <- wordcountaddin::word_count( )
```

Word count: 1250

Word count (alternative): 1445

5.4 Include figures and tables

There are different ways to include figures and tables. I do not recommend doing it with html code as this does not work with pdf output. Below you see two identical pictures of me.

Figure 1: Prof. Dr. Stephan Huber[1]

Figure 2: Prof. Dr. Stephan Huber

Literature

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R markdown: The definitive guide.* Retrieved January 30, 2023; Chapman; Hall/CRC. Retrieved from https://bookdown.org/yihui/rmarkdown/

Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook.* Retrieved January 30, 2023; Chapman; Hall/CRC. Retrieved from https://bookdown.org/yihui/rmarkdown-cookbook