Project Work in Data Science for Business

Dr. Stephan Huber[1] (Module Coordinator)
your name (Student) Michael Jackson (Student)
Bob Ross (Student)

May 9, 2020

[1]stephan.huber@hs-fresenius.de

Preface

This document should help to describes and paraphrases how the project work in the course *Data Science for Business* should look like. The paper is structured as follows: Chapter 1 gives an idea of how the project paper can be structured. In chapter 2 the titles of the projects are listed and chapter 3 offers a brief outlook.

Overall, the idea of the project is that we write a textbook together that can work as an applied introduction to data science for students of business administration. Each project consists of one chapter. Of course, students can work together, form groups, and divide work with respect their preferences and talents, respectively. However, every student is responsible (and is graded) for his/her chapter and his/her presentation only.

Moreover, I will implement (and monitor) your working process using GitHub. I will explain you how that works in a video.

Please notice, that this document is not written to be self-explaining. It should rather work as a template. I explain everything in further detail during a GoToMeeting.

Contents

List of Tables

List of Figures

Chapter 1

A Sketch of How a Project Should Look Like

Barbara Streisand
barbara@email.de

Summary This chapter offers a sketch on how a project may be structured. A project should include a summary, learning objectives, a motivation, a methodolocigal section, a section with an application in R, an example from businesses, a conclusion, exercises, test questions, and a glossary.
An abstract should make clear in simple language what the chapter is all about. Its maximum length is 200 words.

Learning Objectives

- ;weruhkltewiouthb
- swdfgwserg

1.1 Motivation

Motivate your topic. Discuss why the topic and the content of your chapter may be of importance for the reader. Make the content of your chapter subject of a discussion. Outline clearly what the reader can expect. Include a description of contents. For example, write at the end of this section "The remainder of the chapter is structured as follows: Section 1.2 introduces the

theory...". Additionally, mention which related topics will not be discussed. Recommend literature for self-study here.[1]

1.2 Methodological Issues

Bring the statistical theory or econometric approach to the reader. Try to raise a basic understanding of the meaning and the difficulties of the respective method.

1.3 Applications in R

Explain how your topic can be addressed by using R and RStudio, respectively. In particular, introduce R-packages that allow to implement the respective methods. Offer a R-script that allows the reader to understand both the method and the programming skills that are needed to apply the method. It should be possible for the reader to replicate the stuff that is presented in the section and the script.

1.4 Example(s) From the Real World

Make an impressive[2] example from the *real world* to deepen the application and the methods discussed above. The example can stem either from the *academic world* or the *business world*. For example, pick an *academic* research paper and summarize it to the reader in a way that he can understand the strengths and weaknesses of the respective investigation. Or, you can explain a business case where the method of interest plays an important role to solve a problem or to earn money, for example.

Sections 4.3 and 4.4 can be combined.

1.5 Conclusion

Conclude the chapter. This may include a short summary, an outlook and/or related literature.

[1] Making a footnote or citing in LaTeX is easy. For example, Provost and Fawcett (see 2013, p. XY) or (Provost and Fawcett, 2013)

[2] The example should be a significant academic contribution, i.e., the example is published in a highly ranked academic journal. Alternatively, the economic impact of the example should be sizable.

1.6 Exercises

(1) Design an exercise that helps the reader to understand and repeat important concepts.

   a) Challenge the reader to think about the topic.

   b) Encourage a discussion.

   c) Bring up new aspects of the method.

   d) Be creative.

1.7 Test Questions

- Ask five short test questions that refer to five important insights from your chapter. A discerning reader should have no problems with these questions.

1.8 Glossary

word This text should define the meaning of word.

Chapter 2

List of Topics

Michael Jackson
mj@king.com

Summary Read the following titles of potential projects and send me an email with three topics you could imagine working on (`stephan.huber@hs-fresenius.de`). Rank them according to your preferences. I will try to consider your wishes. Feel free to make your own suggestions. But please, don't forget to name overall three (!) topics. If everybody likes to have the same topic, I flip a coin and/or try to consider your second and third preference.

1. Data Collection in Data Science
2. Exploratory Data Analysis in Data Science
3. Regression Analysis in Data Science
4. Spatial Analysis in Data Science
5. Nearest Neighbor Analysis in Data Science
6. Decision Trees Analysis in Data Science
7. Text Mining Analysis in Data Science
8. Big Data Analysis in Data Science
9. Machine Learning in Data Science
10. Single-Board Computers in Data Science

Chapter 3

Further Procedure

Bob Ross
goodnight@paint.org

The project work includes the following:

- Write a chapter on you topic as paraphrased above. Length of the paper should be about 25 pages.
- Prepare a presentation of 20 minutes length (we will see how we manage that part).
- Use the Latex to write the chapter.
- Use Github to include your chapter into the book which contains all chapters.
- Use R and upload your R-code to Github.

3.1 LaTeX

LaTeX is easy and powerful. It allows the author to focus on the content because he don't need to take care about the layout very much. There are millions of sources online that provide tutorials. You can include figures in an floating environment[1], see Figure 3.1. Tables are possible, too. See Table 3.1 or you can reference online sources, see `https://texdoc.net/texmf-dist/doc/latex/booktabs/booktabs.pdf` for a guide to make nice tables.

[1]The placement of the figure is optimized automatically.

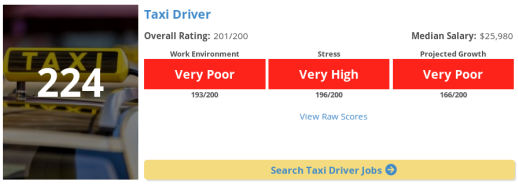

Figure 3.1: This is a picture

Table 3.1: This is a table

Item		
Animal	Description	Price ($)
Gnat	per gram	13.65
	each	0.01
Gnu	stuffed	92.50
Emu	stuffed	33.33
Armadillo	frozen	8.99

3.2 Git and GitHub

GitHub provides hosting for software development version control using Git which is a version control system designed to handle projects with many contributors. Both tools are heavily used in software engineering and data science. It is particularly powerful when teams work on projects with a procedural workflow.

3.3 Style of Writing

Assume your reader is a well informed master student of business administration. Make things easy for your reader. Believe me, this sounds so easy but it is actually the most difficult task for scientists and academic writers. I got a lot out of reading *Writing Tips for Ph. D. Students* from Cochrane (2005)[2]. While you are not a Ph. D. student, these writing tips apply for all authors who aim to communicate efficiently.

[2]You can download this file using `https://t1p.de/br5o`

Chapter 4

Dictionary-based sentiment analysis

Ryan Zidago
zidago.ryan@hs-fresenius.de

Summary This is my abstract.

Learning Objectives

- ;

4.1 Motivation

4.2 Methodological Issues

4.3 Applications in R

4.4 Example(s) From the Real World

4.5 Conclusion

4.6 Exercises

(1) Design

a) C

4.7 Test Questions

- A

4.8 Glossary

w T

Bibliography

Cochrane, John, "Writing tips for PhD Students," Technical Report 2005.

Provost, Foster and Tom Fawcett, *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking*, O'Reilly Media, 2013.