EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Wirtschafts- und Sozialwissenschaftliche Fakultät

LEHRSTUHL FÜR MARKETING
Prof. Dr. Stefan Mayer

Telefon +49 7071 29-76979
stefan.mayer@uni-tuebingen.de
www.uni-tuebingen.de/wiwi/marketing

Data Science Project – Summer Term 2025

As part of the M.Sc. „Data Science in Business and Economics"

1 Introduction

All students in the master program M.Sc. Data Science in Business and Economics enrol in an implementation-focussed module "Data Science Project". With 12 ECTS, the project has a substantial weight in the curriculum. This is also reflected in our expectation that students show a high degree of independence in their work in this module. Students work together in groups of 2-4 persons.

The basic idea of this module is that students **independently** design and program the entire process of a data science project from start to finish.

This requirement places a focus on **automation**, which means that the procurement of data, the structuring, reading, validation, modification, and analysis take place (as far as possible) without the intervention of the analyst or user of the analysis.

A data science project should include the following components:

1. **Data acquisition**: The project acquires and reads in the data, for example, by scraping data from web pages, by using one or several APIs, or by reading text documents, images, or video data. It is also possible for students to create a new dataset semi-manually themselves or to use an existing dataset as the basis for the project. The complexity and the effort of the data acquisition are considered in the grading of the project.
2. **Data preparation and validation**: The data are read in and made ready for further analysis, computation, and/or estimation. Among other things, this means that unstructured data (e.g., texts) are converted into computable formats and that different data sources are combined meaningfully. The data should be validated, for example, by generating descriptive statistics and figures and by identifying implausible observations and outliers. After having defined the criteria for validation, this process should also take place mainly without the users' intervention.
3. **Data analysis**: The primary focus of the project is a relevant substantive question, which can, but does not necessarily have to be of an economic nature. The

corresponding analysis should be implemented with adequate methods (traditional statistics or machine learning).

4. **Validation of the results**: It is crucial to validate the results of the analysis by suitable approaches. This is to make sure that the results from your analysis are helpful and practically relevant.
5. **Presentation of the results**: All project outputs (e.g., tables, graphs, results) should be embedded in an interactive environment, for example, in the form of a dashboard. Results should be presented in an appealing and informative way, according to the state-of-the-art of data visualization.
6. **Automation**: The entire project with the above-mentioned steps should run automatically. When new data periodically arrives, a user who did not create the code can evaluate and analyse it without having to "touch" the code.
7. **Version control and collaboration:** As soon as the topic, data sources, structure, and type of analysis are defined, the code, including the version history, must be organized in a GitHub repository.
8. **Documentation**: The whole project should be well-documented so that external parties can easily understand and review what is happening in your code and how different components of your architecture work together. This includes a meaningful structure of your repo (directory structure and file names) and comprehensibly commented code, among others. In addition, there should be a document that guides the reader through your repo (e.g., a README) that includes, for example, which files must be executed and in what order.

2 What is the difference between a master thesis and the Data Science Project?

A master thesis focuses on solving a relevant research problem, which must be embedded in the relevant literature. The analyses are described in the text, and the results are printed in the form of static tables or graphs. The thesis then interprets the results and draws conclusions. There are no requirements for automation, and the reader "consumes" the text without interacting with the data themselves.

	Master thesis	**Data Science Project**
Elaborate research problem	High weight	Low weight
Literature work	High weight	Low weight
Automation	Low weight	Very high weight
Reader/user interacts with data	Low weight	Very high weight
Conclusion	Mostly author	Rather reader/ user

3 Evaluation/ Grading

Students present their results (as a group) twice: In an interim presentation, they present the plan and first steps. In the final presentation, they show the completed project, the central structure of the code, the architecture behind their “product”, and how users can interact with the result.

The project is evaluated according to the following criteria:

1. Complexity of the problem
    a. How complex are the different components of the project?
    b. How novel is the problem and the proposed solution?
2. Quality of the execution
    a. How is the user experience of the final product?
    b. How well does the automation work? How well is the error-handling?
    c. Is the implementation logically consistent?
3. Quality of the documentation
    a. Are the project architecture and code structure easily understandable?
    b. Is the code easily readable by an external party?

For assessment, students will produce the following components:

1. A presentation that motivates the central question, derives its relevance, and refers to relevant literature as appropriate. Furthermore, students present their project’s architecture and technical implementation as well as the finished “product”, e.g., a website/dashboard or a comparable way to interact with the data as a user. The presentation also contains a discussion in which supervisors and other students ask questions.
2. Students provide the code by granting the supervisors access to the GitHub repository or by providing a link to the repository if it is public.
3. Since the data set might be very large, there is no need to upload it to GitHub. However, be prepared to grant supervisors access, e.g., by a dedicated database user.
4. Students make the interface (e.g., Dashboard) available to other students and teachers, e.g., by hosting it on bwCloud.
5. Students create a short video (approx. 3 minutes) that clearly presents the project and the result to a broader circle of users. The video can be a screencast where only the presenter's voice is heard while the final product is shown, or it can be a video presentation in which the presenters are also visible. The target group is a non-technical audience; e.g., in a business context, it would be a manager, not a fellow data scientist.

Evaluation is done at the individual level, not at the group level. Thus, the group members’ contributions to the project and presentations will be considered in the evaluation.

All components (link to code, video, link to interface) are provided to the supervisor one week prior to the final presentation (see schedule below).

4 Schedule

Monday, April 7, 2025 13:00 s.t. – 14:00 s.t.	Kick-off, assignment of topics, supervisors, and teams	SR 331, Mohlstr. 36
Tuesday, April 8, 2025 9:00 s.t. – 13:00 s.t.	Workshop “Presentation Skills”	SR E03, Mohlstr. 36
Friday, May 23, 2025 9:00 s.t. – 13:00 s.t.	Interim presentation	SR 281, OSA Keplerstr.
Friday, July 11, 2025 23:59	Submission of project (except slides for the final presentation)	---
Thursday, July 17, 2025 20:00	Submission of slides for the final presentation	---
Friday, July 18, 2025 9:00 s.t. – 13:00 s.t.	Final presentation	FSR, Nauklerstr. 47

Note: Participation on all dates is mandatory. Times might change depending on the number of participants.

5 Topics

We encourage students to contribute their own topics, so every student presents a topic suggestion at the kick-off. Specifically, students prepare a one-minute pitch that briefly explains their suggested topic, including the problem to be solved, its relevance, potential approaches, and data sources. To inspire students, we will publish the videos (see above) from previous data science projects on ILIAS.

We will also present some topics at the kick-off, so students are not forced to choose the topics they present themselves.