

Project2023

Team:

- *) Ellar Seidelberg**
- *) Free place**
- *) Free place**

Project link:

(<https://github.com/EllarSeidelberg/Project2023>)

Report:

Task 1. Setting up

- business understanding
- data understanding
- planning

Kasutame <https://avaandmed.eesti.ee/> andmebaasi.

(ei tea veel millist).

Aga tunduvad head andmebaasid selles mõttes, et on CSV formaadis ja kergesti kättesaadavad – seega pole andmete kogumisega probleeme.

Task 2. Business understanding

NB! Don't forget to mention your project title and team members at the beginning of the report.

Developing a business understanding within CRISP-DM consists of four tasks: identifying your business goals, assessing your situation, defining your data-analysis, data-mining or machine learning goals and producing your project plan. For this exercise, please develop a business understanding of your project. According to CRISP-DM, you should report the following:

- Identifying your business goals
 - Background
 - Business goals
 - Business success criteria
- Assessing your situation
 - Inventory of resources

- Requirements, assumptions, and constraints
 - Risks and contingencies
 - Terminology
 - Costs and benefits
- Defining your data-mining goals
 - Data-mining goals
 - Data-mining success criteria

Please follow this given structure and cover all these aspects in your report. Consult [this PDF-file with a chapter on Embracing the Data-Mining Process for more information on each of the deliverables](#). Keep the report concise, and feel free to state that some aspect is irrelevant to your project. If your project is not meant to benefit a ‘business’, then please specify who will benefit from the project and perform business understanding from their perspective. For instance, this could be either one or multiple individuals, organizations, or societies. Please focus on the goals that you plan to directly contribute to, not on the generic goals (like making the world a better place).

The report of task 2 should be 400-800 words.

Task 3. Data understanding

Data understanding within CRISP-DM consists of performing four tasks: gathering data, describing data, exploring data and verifying data quality. For this exercise, please develop a data understanding of your project. Report the results of the tasks according to the following structure:

- Gathering data
 - Outline data requirements
 - Verify data availability
 - Define selection criteria
- Describing data
- Exploring data
- Verifying data quality

Consult the above-given book chapter to understand what is expected under all these deliverables. Take inspiration from when describing and exploring the data. As a result of this exercise, you should have gathered and understood the data. You should have decided which parts of the data you are potentially going to use and understood the meaning of all fields within these parts. Note that data cleaning is part of the data preparation step in CRISP-DM, but you might choose to do some of it already during this task.

The report of task 3 should be 400-800 words.

Task 4. Planning your project

Please perform the following tasks:

- Make a detailed plan of your project with a list of tasks. There should be at least 5 tasks. Specify how many hours each team member is going to contribute to each task.
- List the methods and tools that you plan to use. Add any comments about the tasks that you think are important to clarify.

Project plan:

*) Task1

*) Task2

*) Task3

*) Task4

*) Task5

The report of task 4 should be 100-300 words.